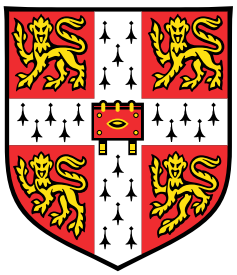
Vision-based trailer pose estimation for articulated vehicles



Christopher Charles de Saxe

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Jesus College September 2017

To Rachel.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains 64,594 words including appendices, bibliography, footnotes, tables and equations and has 114 figures.

Christopher Charles de Saxe
September 2017

Acknowledgements

I would like to acknowledge the following individuals, groups and organisations:

- Professor David Cebon for his expert guidance and unwavering confidence throughout this project.
- The Cambridge Commonwealth, European and International Trust (CCEIT) and the Council for Scientific and Industrial Research (CSIR) for funding for this work through the CSIR South Africa Cambridge Scholarship.
- The W G Collins Endowment Fund, Ford of Britain Fund, and Jesus College Cambridge for contributing towards conference expenses.
- Member organisations of the Cambridge Vehicle Dynamics Consortium (CVDC) for general funding to the research group which has supported various equipment and travel expenses associated with this project.
- Johan Eklöv and Gustav Neander of Volvo Group Trucks Technology in Gothenburg, for their interest in my work and for support with vehicle testing at Hällered.
- Dr Ankur Handa (previously at the University of Cambridge) and Dr Pedro Pinies (University of Oxford) who contributed valuable suggestions relating to computer vision methods.
- The Cambridge University Engineering Department technicians, particularly Peter Down, for workshop and vehicle testing support.
- Current and previous members of the Transportation Research Group—Dr Richard Roebuck, Dr Will Midgley, Dr Leon Henderson, Dr Amy Rimmer, Dr Yanbo Jia, Matthew Harris, Dr Terry Xu, Dr Matthew Miao, Dr Graeme Morrison, Dr Xiaoxiang Na, Ayelet Ashkenazi, Chris Eddy, Albert Liu, Francesco Amoruso, Dr Julien Lepine, and Virginia Longhi—for ongoing technical advice, vehicle testing support, good humour, and daily tea breaks and crosswords.

- The friends I've made in Cambridge, who have taught me so much about the world, and who have given me many special memories of my time here.
- My family, Patricia, Bruce, and Paula, for having the faith in me to achieve my goals and for always supporting and encouraging me; my late grandfather Professor Ian Watt, for being my first engineering inspiration; and also to the Cameron family for their ongoing support.
- My loving wife, Rachel, who has selflessly followed me as I've pursued my goals, and who has provided absolute support and encouragement throughout every high and low of the past four years.

Summary

Articulated Heavy Goods Vehicles (HGVs) are more efficient than conventional rigid lorries, but exhibit reduced low-speed manoeuvrability and high-speed stability. Technologies such as autonomous reversing and path-following trailer steering can mitigate this, but practical limitations of the available sensing technologies restrict their commercialisation potential. This dissertation describes the development of practical vision-based articulation angle and trailer off-tracking sensing for HGVs.

Chapter 1 provides a background and literature review, covering important vehicle technologies, existing commercial and experimental sensors for articulation angle and off-tracking measurement, and relevant vision-based technologies. This is followed by an introduction to pertinent computer vision theory and terminology in Chapter 2.

Chapter 3 describes the development and simulation-based assessment of an articulation angle sensing concept. It utilises a rear-facing camera mounted behind the truck or tractor, and one of two proposed image processing methods: template-matching and Parallel Tracking and Mapping (PTAM). The PTAM-based method was shown to be the more accurate and versatile method in full-scale vehicle tests. RMS measurement errors of 0.4–1.6° were observed in tests on a tractor semi-trailer (Chapter 4), and 0.8–2.4° in tests on a Nordic combination with two articulation points (Chapter 5). The system requires no truck-trailer communication links or artificial markers, and is compatible with multiple trailer shapes, but was found to have increasing errors at higher articulation angles.

Chapter 6 describes the development and simulation-based assessment of a trailer off-tracking sensing concept, which utilises a trailer-mounted stereo camera pair and visual odometry. The concept was evaluated in full-scale tests on a tractor semi-trailer combination in which camera location and stereo baseline were varied, presented in Chapter 7. RMS measurement errors of 0.11–0.13 m were obtained in some tests, but a sensitivity to camera alignment was discovered in others which negatively affected results. A very stiff stereo camera mount with a sub-0.5 m baseline is suggested for future experiments.

A summary of the main conclusions, a review of the objectives, and recommendations for future work are given in Chapter 8. Recommendations include further refinement of both sensors, an investigation into lighting sensitivity, and alternative applications of the sensors.

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

a	Longitudinal distance from fifth wheel to stereo cameras
a_1, a_2	CoG location behind front axle/hitch point (tractor, trailer)
a_{RT}	Longitudinal distance from fifth wheel to RT3022
b	Longitudinal distance from stereo cameras to trailer follow-point
b_1, b_2	CoG location ahead of rear axle/axle group (tractor, trailer)
b_{RT}	Longitudinal distance from RT3022 to trailer follow-point
c	Lateral offset of stereo cameras from trailer longitudinal axis
c_1	Hitch location forward of rear axle/axle group
C_f, C_r, C_t	Tyre cornering stiffness (front, rear, trailer)
c_{RT}	Lateral offset of RT3022 from trailer longitudinal axis
d	Distance from camera to trailer face
e	Incremental lateral motion of the centre of the trailer face
e'	Incremental lateral pixel motion of the centre of the trailer face
$\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$	Vector components of R_{rect}
e_{tr}	Trailer off-tracking, measured laterally at trailer follow-point
$\mathbf{F}$	Process model
f	Focal length

F_f, F_r, F_t	Lateral tyre forces (front, rear, trailer)
F_h	Hitch force
FOV	Field of view
f_x,f_y	Focal length (horizontal, vertical)
g	Trailer half-width
$\mathbf{H}$	Measurement model
H	Height of optical axis above trailer hitch point
h	Trailer front overhang
I	Image
I_t	Template image
$\bar{\mathrm{I}}_t$	Template image mean
$\bar{\mathrm{I}}_{u,v}$	Image mean under the area overlapped by I_t
J_1, J_2	Yaw moment of inertia about the CoG (tractor, trailer)
K	Camera calibration matrix
k_1, k_2, ...	Radial distortion coefficients
k_b	Length of data buffer
k_u, k_v	Pixel scaling parameters for rectangular pixels
L	Length of random variable $\mathbf{x}$
l_1, l_2	Wheelbase (tractor, trailer)
l_{eq}	Equivalent wheelbase
m_1, m_2	Vehicle mass (tractor, trailer)
m_u, m_v	Template image width and height
$\mathbf{N}$	Normal vector to trailer front plane at datum
$\mathbf{n}$	Measurement noise

n_a	Number of axles in an axle group
n_u, n_v	Image width and height
$\mathbf{P}$	Covariance vector
p_1, p_2	Tangential distortion coefficients
P	Planar homography matrix
$\mathbf{Q}$	Process noise covariance
R	Rotation matrix
r	Radial image coordinate
R_{11}, R_{12}, ...	Rotation matrix components
r_{PTAM}	Radial image coordinate (PTAM)
R_{rect}	Stereo rectification rotation matrix
R_x, R_y, R_z	Roll, yaw, pitch rotation matrices
s	Distortion parameter (PTAM)
$\mathbf{T}$	Translation vector
TF	Tandem factor
T_x, T_y, T_z	Translation vector components
u, v	Pixel coordinates (horizontal, vertical)
u_0, v_0	Principal point coordinates (horizontal, vertical)
U_1, U_2	Longitudinal velocity (tractor, trailer)
$\mathbf{v}$	Process noise
V_1, V_2	Lateral velocity (tractor, trailer)
$\mathbf{w}$	Pixel coordinates, $\mathbf{w} = (u, v)$
$\mathbf{W}$	Measurement noise covariance
$W^{(c)}$	Sigma point weight, w.r.t. variable covariance (UKF)

$W^{(m)}$	Sigma point weight, w.r.t. variable mean (UKF)
$\mathbf{X}$	World coordinates, $\mathbf{X} = (X, Y, Z)$
$\mathcal{X}$	Sigma point (UKF)
x, y	General location coordinates in the horizontal plane
$\mathbf{x}$	Gaussian random variable
$\bar{\mathbf{x}}$	Mean of Gaussian random variable $\mathbf{x}$
$\hat{\mathbf{x}}$	Estimated mean of Gaussian random variable $\mathbf{x}$
$\mathbf{X}_c$	Camera-centred coordinates, $\mathbf{X}_c = (X_c, Y_c, Z_c)$
x_{cam}, y_{cam}	Location coordinates of the stereo cameras
x_F, y_F	Location coordinates of the trailer fifth wheel
$\mathbf{X}_I$	Image plane coordinates, $\mathbf{X}_I = (X_I, Y_I, Z_I)$
x_R, y_R	Location coordinates of the trailer follow-point
$Z_{c,A}$	Z_c at the centre of the trailer face

Greek Symbols

α	Sigma point spread (UKF)
α_f, α_r, α_t	Tyre slip angle (front, rear, trailer)
β	Prior knowledge parameter (UKF)
Γ	Articulation angle
Γ_{lim}	Γ at the point where the trailer face is perpendicular to the camera line-of-sight
Γ_{trans}	Γ at the point where lateral pixel motion of the trailer front and side are equal
γ	Correlation coefficient
δ	Steer angle
ε	Error
θ	Pitch angle

κ	Scaling factor (UKF)
λ	Scaling parameter (UKF)
ξ	Arbitrary scale factor for homogeneous co-ordinates
ρ	Axle spacing
ϕ	Roll angle
Ψ	Heading angle
ψ	Yaw angle
ψ_{c2v}	Yaw rotation from stereo cameras to vehicle reference frame
Ω_1, Ω_2	Yaw rate (tractor, trailer)

Acronyms / Abbreviations

ABS	Anti-lock Braking System
CAD	Computer-Aided Design/Drawing
CAN	Controller Area Network (also 'CANbus')
CCD	Charge-Coupled Device
CNN	Convolutional Neural Network
CoG	Centre of Gravity
CVDC	Cambridge Vehicle Dynamics Consortium
DC-GWNS	Dual Camera Ground Watching Navigation System
DfT	Department for Transport (UK)
DoG	Difference of Gaussian
EBD	Electronic Brakeforce Distribution
EBS	Electronic Braking System
EKF	Extended Kalman Filter
EMS	European Modular System (of EU Directive 96/53 EC)

ESC	Electronic Stability Control
FAST	Features from Accelerated Segment Test
FLANN	Fast Library for Approximate Nearest Neighbours
GCM	Gross Combination Mass
GLOH	Gradient Location and Orientation Histogram
GPS	Global Positioning System
HCV	High Capacity Vehicle
HGV	Heavy Goods Vehicle
HOG	Histogram of Oriented Gradients
IMU	Inertial Measurement Unit
LCV	Long Combination Vehicle
LED	Light-Emitting Diode
LIDAR	LIght Detection And Ranging
LoG	Laplacian of Gaussian
LRR	Long Range Radar
LST	Longer Semi-Trailer
NCC	Normalised Cross-Correlation
ODE	Ordinary Differential Equation
OEM	Original Equipment Manufacturer
PBS	Performance-Based Standards
PTAM	Parallel Tracking and Mapping
RADAR	RAdio Detection And Ranging
RANSAC	RANdom SAmple Consensus
RGB	Red Green Blue (pixel colours)

RMS	Root Mean Squared
RTK-GPS	Real-Time Kinematic GPS
SC-GWNS	Single Camera Ground Watching Navigation System
SIFT	Scale-Invariant Feature Transform
SLAM	Simultaneous Localisation and Mapping
SRR	Short Range Radar
SURF	Speeded Up Robust Features
TM+UKF	Template-matching algorithm with UKF
TM	Template-matching (algorithm)
UAV	Unmanned Aerial Vehicle
UKF	Unscented Kalman Filter
VGA	Video Graphics Array (640×480 image resolution)
VISO-OT	Combined VISO2-S (visual odometry) and off-tracking algorithm
vSLAM	Visual Simultaneous Localisation and Mapping

Chapter 1

Introduction and literature review

1.1 Background and motivation

Heavy Goods Vehicles, or HGVs, play a vital role in sustaining and growing modern economies [6]. In the UK in 2014, inland freight transport amounted to 185 billion tonne kilometres, and of this 74% was transported by road on HGVs [7]. This freight movement serves all aspects of society, with 23% of 2015 HGV traffic transporting food products, followed by 'metal, mineral and chemical products' (18%) and 'products of agriculture, forestry, raw materials' (18%) [7]. Other important commodities transported by HGVs include machinery and equipment, consumer goods, waste and mail.

The growth of freight transport demand on limited infrastructure, coupled with ambitious CO_2 emission reduction targets, has driven technology and policy development to investigate and implement measures to improve the efficiency of road freight transport. The latest developments in engine technology yield vehicle CO_2 reductions in the region of 1–2% [8], at increasing costs and diminishing returns as internal combustion engine technology saturates. Low rolling-resistance tyres and improved aerodynamics are relatively low cost measures with few barriers to adoption, offering reductions of 1–5% and 6–9% respectively for long haul applications [9].

Alternative fuel technology such as methane can reduce long haul CO_2 emissions by 0–12%, but presents significant barriers in terms of fuelling infrastructure and associated cost [9]. Hybrid-electric engine technology offers long haul reductions of 4–10% and does not require any additional infrastructure, but can significantly increase vehicle cost [8].

Full electric powertrains reduce 'tank-to-wheel' carbon emissions to zero, but their realistic 'well-to-wheel' emissions are directly linked with the carbon generated by the power grid. At present, full electric vehicles present an emissions reduction of approximately 40% relative to conventional diesel vehicles [8], but this figure will improve with further

decarbonisation of the power grid. However, full electric powertrains are currently limited to vehicles up to 12 tonnes gross mass and so are not suited to long haul applications, and introduce significant barriers at present in terms of supporting infrastructure and costs [8].

In comparison, it is possible to reduce vehicle emissions by 11–19% [10] (and higher in certain cases [11]) with relatively minor barriers to introduction by permitting the use of longer and/or heavier truck combinations, carrying more freight with fewer vehicles. These vehicles are often referred to as 'High Capacity Vehicles', or HCVs. This solution can utilise existing trucks and trailers, requires no additional infrastructure, and reduces the number of HGVs on the roads. Furthermore, these vehicles are as safe and often safer than conventional HGVs when operated within a suitable regulatory framework [11]. The effectiveness and safety of such vehicles has been proved in implementations and trials in a number of countries [12].

The barriers for incorporating the use of HCVs on arterial roads and motorways ('primary routes') are mostly political, social, and regulatory. Successful trials and implementation of HCVs in Australia, New Zealand, Canada, South Africa, and parts of Europe [13–17] have demonstrated that these barriers can be successfully overcome, and that HCVs are a feasible and effective solution.

Provided HCVs undergo an informed design process ensuring inherent dynamic safety, and are operated within a suitable regulatory framework ensuring best practice in terms of loading, driver training, and maintenance, their adoption on primary routes requires little in the way of technological intervention. However, the net emissions reduction benefit of HCVs is clearly a function of uptake. Improving the accessibility of HCVs to secondary road networks would enable them to serve a wider spectrum of transport tasks, thereby increasing uptake relative to conventional HGVs. Technological interventions can help to facilitate this.

In some cases technological intervention would be mandatory in order to meet country-specific legislation. In the UK for example, all HGVs must meet the turning circle requirement [18], which limits the level of 'cut-in' as a vehicle navigates a standardised roundabout manoeuvre. Long HCVs have inherently reduced manoeuvrability and are unlikely to meet this requirement without trailer steering technology.

Manoeuvrability is a particular limitation for long HGVs or HCVs with two or more articulation points. These will tend to exhibit large cut-in behaviour when turning, and are inherently unstable in reverse [19]. Although a professional driver can reverse singly-articulated vehicles (into a loading bay for example), reversing combinations with two or more articulation points is nearly impossible without technological intervention.

Technologies such as active trailer steering [20–22] and autonomous reversing [23] have been shown to improve the manoeuvrability and hence road network access of HCVs.

Similarly, technologies such as combined braking and steering control [24] and anti-jackknife control [25] help to improve the high-speed safety of HCVs, improving their motorway performance in a number of conditions. Many of these technologies have been developed within the Cambridge Vehicle Dynamics Consortium (CVDC).

Commercialisation and wide-spread uptake of these technologies requires robustness and financial feasibility in order to be attractive to fleet operators and original equipment manufacturers (OEMs). Enforcing their adoption through regulation is possible, but legislators are wary of the economic effects on an industry which already operates with very low profit margins [26].

A notable challenge preventing the widespread uptake of such technologies is the requirement for additional vehicle sensors and instrumentation. Specialised sensing equipment is fundamental to most active vehicle control technologies (see for example [20, 27, 28]), and high levels of accuracy and robustness are usually required. The cost and practicality of sensors is equally important.

Autonomous reversing, jackknife control, combined braking and steering control, and trailer steering are important technologies in the effort to enable the use of HCVs in more countries, and on wider subsets of the road network. Two sensing requirements have been identified as particular barriers to the commercial adoption of these technologies: articulation angle sensing and trailer off-tracking estimation.

Articulation angle sensing refers to the measurement of the yaw angle between a truck and trailer, or between subsequent trailers, and is a core requirement for autonomous reversing, jackknife prevention, and combined braking and steering control technologies [29–31]. Assuming conventional trailers, it is necessary for articulation angle sensing systems to be based on the tractor unit, where control processing and actuation signals originate. Currently however, these technologies rely on either trailer-based articulation angle sensors with non-standard or experimental communication links[1], or estimation techniques which require knowledge of the trailer states and other non-standard sensors.

Trailer off-tracking refers to the lateral deviation of the rear of a trailer from the path followed by the tractor unit. This can either originate from the familiar ‘cut-in’ behaviour a trailer exhibits as an articulated vehicle navigates a turn, or when an articulated vehicle is travelling on a low friction surface with cross-slope. Off-tracking estimation is fundamental to the ‘path-following’ trailer steering technology developed within the CVDC [32, 20]. Trailer-steering is a trailer-based control technology, but off-tracking is currently determined

[1] Conventional trailers have limited communication links between the truck/tractor and trailer. Currently the ABS or EBS CANbus link for braking signals is the only standard data communication link between the two.

through a combination of tractor and trailer sensors and knowledge of geometric properties including those of the tractor. It also assumes high levels of friction at all wheels.

There is an important distinction to make at this point, regarding tractor- and trailer-based control technologies, and their ideal sensing requirements. The distinction is partly due to the inherent nature of the HGV industry. Throughout the world, tractor and trailer units are designed to be interchangeable so that any tractor can pull any trailer. This gives maximum flexibility in logistics operations. Furthermore, truck manufacturers produce trucks and develop truck-based technology, while trailer manufacturers produce trailers and develop trailer-based technology. Because of the interchangeability requirement, tractor-based technologies should require no knowledge of or modifications to a trailer, and vice versa.

If a tractor-based technology such as autonomous reversing requires the fitment of specialised sensors or modifications to a trailer, it is limited to function only with that trailer. In practice, one tractor unit may be used for various freight tasks, and coupled to a variety of trailers. Trucks and trailers may also be coupled and decoupled multiple times within a single day or transport task, and this is especially true in the context of the European Modular System (EMS) (more on this later). Therefore, all trailers which could be used with that tractor would need to be fitted with the sensors, or modified in some other way, adding cost and burden to the operator. These burdens on the operator form a barrier to the adoption of the technology, and hence the adoption of HCVs in general.

A substantial improvement in the commercial feasibility of these technologies would be achieved if all sensing could be conducted remotely on the same vehicle unit (*i.e.* tractor/trailer) as the controller and actuation, without the need for modifications to or pre-existing knowledge of the other vehicle unit. Furthermore, this would permit these technologies to be completely developed, optimised, and pre-fitted by the respective OEM, mitigating the risk of incorrect fitment of sensors or modifications to the other vehicle unit which would impact performance.

The ideal scenario of purely tractor-based control for autonomous reversing is illustrated in Figure 1.1a. In this case the controller is tractor-based, as is actuation via the steered wheels, and so ideally sensing should also be tractor-based. Similarly for jackknife control (Figure 1.1b), actuation is in the form of either steer input, trailer brake actuation, or hitch locking/damping, and so the system can be entirely tractor-based. Trailer brake actuation signals can be sent over standard EBS[2] CANbus[3] links between tractor and trailer, mandatory on all articulated HGVs in Europe.

[2]Electronic Braking System

[3]Controller Area Network (data bus standard for vehicles)

The ideal scenario for combined braking and steering control is shown in Figure 1.1c. Again actuation includes the trailer brakes but actuation signals can be sent over standard CANbus connections[4]. Ideal trailer-based trailer steering control with trailer-based off-tracking estimation is shown in Figure 1.1d.

The tasks of articulation angle and trailer off-tracking measurement are fundamentally 'pose estimation' and 'ego-motion' estimation problems. Pose estimation pertains to determining the position and orientation of a body (*i.e.* the trailer angle) from an external reference (the truck/tractor). Ego-motion estimation is the problem of determining a body's motion (*i.e.* trailer off-tracking) from the reference of the observer (the trailer).

In summary, in order to facilitate the wider uptake of HCVs, there is a need to address the barriers to their implementation. One of the important barriers is the commercial feasibility of enabling technologies such as autonomous reversing and path-following trailer steering. Two particular sensing requirements have been identified as barriers to the adoption of these technologies, and this research sets out to investigate potential solutions to these needs. These can be summarised as the need for:

1. tractor-based articulation angle sensing, and
2. trailer-based, trailer off-tracking estimation[5].

These sensing systems should not require modifications to the other vehicle unit (tractor or trailer), should be low cost and practical, should be sufficiently accurate for control effectiveness, and should be robust to the heavy duty operating environment of HGVs. Non-contact sensors would be preferable to contact sensors to ensure durability under the heavy loads and industrial operating environments of HGVs.

The following section provides a more detailed introduction to HGVs and associated terminology, followed by a section on HCVs and their use in various countries. Thereafter, a review of pertinent 'enabling technologies' for HCVs is given, highlighting current limitations. A review of existing commercial and experimental sensors for articulation angle and off-tracking estimation is then given. Lastly, various pose and ego-motion estimation methods are discussed, with a focus on current vehicle technologies and vision-based sensing.

[4]Although combined steering and braking control as presented by Morrison [24] requires currently non-standard brake valves, it is fair to assume that the next generation of braking technology will incorporate such fast-acting valves as standard, given their significant impact on braking and steering performance.

[5]Trailer off-tracking measurement from the tractor is theoretically possible, given tractor side-slip and articulation angle measurements. However, trailer-steering actuation remains trailer-based, and so actuation signals would need to be sent from tractor to trailer. This is not possible with current standard tractor-trailer communication links. It is possible that additional communication links may be introduced in the future, necessitated by significant developments in advanced full-vehicle control systems. However, the time scale of these developments was assumed to be beyond that of the current work.

1.2 Heavy Goods Vehicles

The definition of a Heavy Goods Vehicle (HGV) differs between countries, but usually refers to road freight vehicles (*i.e.* trucks or lorries) above a certain Gross Combination Mass (GCM). In the UK, vehicles and vehicle combinations above 3.5 tonnes are considered HGVs. The permitted truck and trailer combinations, maximum GCM, and maximum length of HGVs also varies between countries.

Generally speaking, an HGV consists of either a single rigid truck, supporting its own payload, or a vehicle combination consisting of a prime mover and one or more trailers (see Figure 1.2). The prime mover may be a rigid truck fitted with a pintle hitch coupling at its rear to which drawbar-type trailers can be coupled. Alternatively the prime mover may be a truck-tractor (hereon referred to simply as a 'tractor'), which bears no payload of its own, but is fitted with a 'fifth wheel' coupling via which a semi-trailer can be attached.

Types of trailer shown in Figure 1.2 include semi-trailers, B-link trailers, full-trailers, and drawbar trailers (also known as dog trailers or centre-axle trailers). A semi-trailer is a trailer with no front axles which couples via a fifth wheel and kingpin. A B-link trailer is a semi-trailer with an extended chassis and fifth wheel at its rear for coupling another semi-trailer or B-link trailer. A converter dolly can be used to couple a semi-trailer to a rigid truck or another trailer via a pintle hitch. A full-trailer is conceptually a dolly and semi-trailer combination, except that the fifth wheel and kingpin coupling is replaced by a turntable, creating a standalone articulated trailer. A drawbar trailer has axles located under the centre of the loading area, and a rigid drawbar extending at its front for coupling via a pintle hitch.

In the UK, permitted HGVs include rigid trucks up to 12 m in length and 32 tonnes GCM (Figure 1.2a), tractor semi-trailer combinations up to 16.5 m and 44 tonnes with a maximum semi-trailer length of 13.6 m (Figure 1.2b), and truck and drawbar trailer combinations up to 18.75 m and 44 tonnes [33] (Figure 1.2c). A field trial is currently underway with 'Longer Semi-Trailer' (LST) combinations with semi-trailer lengths of 15.65 m, giving a total combination length of 18.75 m [7]. Conventional HGVs used in other countries include a 'B-double' (Figure 1.2f), common in Australia and South Africa, and the 'Nordic combination' (Figure 1.2e), common in Sweden and Finland up to 25.25 m in length[6].

In the UK the term 'articulated vehicle' normally refers only to tractor semi-trailer combinations, but the term will be used more generally here to refer to any truck and trailer combination, as distinct from a rigid truck.

[6]The 'Nordic combination' (Figure 1.2e) and the 'truck and full-trailer' (Figure 1.2d) combination are conceptually identical, in that they consist of a rigid truck, dolly and semi-trailer. In the case of the full-trailer, the dolly and semi-trailer are typically integrated as one trailer through a 'turntable' coupling. In the case of the Nordic combination, a conventional semi-trailer is 'converted' into a full-trailer via a 'converter dolly' (connected via a fifth wheel coupling), resulting in greater modularity and flexibility.

1.3 High Capacity Vehicles

Traditionally, the cost of logistics has been a primary driver of the transport sector towards improving the efficiency of transport. Another major driver has emerged in the form of ambitious CO_2 emissions targets. The European Union (EU) has set a target of 20% greenhouse gas emission reduction by 2020 (from 1990 levels) and 80% reduction by 2050. The UK has set a target of 35% by 2020 [34].

Domestic transport accounts for approximately 20% of the UK's total CO_2 emissions, and HGVs account for 21% of this, or 4.2% of the total [35]. The case is similar in Europe, with road transport responsible for about 20% of total CO_2 emissions, approximately one quarter of which is attributable to trucks and buses [36]. Therefore, improving the efficiency of road freight transport, and specifically that of HGVs, is important in the efforts to achieve global emission reduction targets.

Enabling the use of longer and/or heavier HGVs, and thereby requiring fewer HGVs for a given freight task, has been shown to be one of the most promising and impactful measures to improve the efficiency of road freight transport, and thereby reduce emissions [12]. It has also been shown that such vehicles can be as safe as or even safer than conventional HGVs, especially when operated under a suitably robust operating framework coupled with self-regulation [12].

HGVs that fall outside of the legislated maximum gross combination mass, maximum length or number of trailers are referred to variously as 'Longer Heavier Vehicles' (LHVs) or 'High Capacity Vehicles' (HCVs). The distinction between HGVs and HCVs is specific to the geographical context however, as a B-double would be considered an HCV in the UK for example, but as a conventional HGV in South Africa where they have operated since the 1980s [37].

The term 'Long Combination Vehicle' (LCV) generally refers to vehicle combinations with two or more trailers. In this case a B-double is an LCV regardless of the context. Most HCVs are LCVs, but exceptions include longer and or heavier tractor semi-trailer configurations. The UK's LST combinations under trial are a good example of this.

In Australia, the use of HCVs has been formally adopted into national transport legislation through the Performance-Based Standards (PBS) scheme [13]. The scheme permits the use of HCVs on designated subsets of the road network, subject to a comprehensive simulation-based assessment of the vehicle dynamics performance, an assessment of the road wear impact and bridge loading effects of the vehicle, special accreditation of the operator, smart access control and tracking of the vehicles, and special requirements for loading control and driver training.

In South Africa, an HCV trial known locally as the 'Smart Truck' scheme [38] has been in operation since 2008, largely based on the Australian PBS scheme. As of February 2017, the trial includes 215 participating vehicles, transporting commodities such as mining ore, timber, fuel, coal and sugar, with vehicles ranging in length from 22 to 40 m, and in mass from 56 to 148 tonnes. To date, performance data for 92.4 million truck kilometres have been accumulated, together with data from conventional vehicles performing the same freight task on identical routes (the 'baseline' vehicles).

The latest data [11] show that the South African Smart Truck fleet saves approximately 67,000 truck trips, 9 million km of truck travel, 2 million litres of fuel and 5,500 tonnes of CO_2 per year. Year-on-year since inception, the Smart Trucks have had between a third and a half of the crash rate of the baseline vehicles, indicating significant safety benefits of the system of vehicles and the Smart Truck framework. The initial trial is expected to conclude in the next year, and discussions towards national implementation are underway.

In Europe in 1996, EU Directive 96/53 EC [39] introduced the European Modular System (EMS). This directive gives EU member states the option to permit HCVs on certain routes within their territory, provided the combinations consist of defined interchangeable vehicle units (as shown in Figures 1.2e, 1.2f, 1.2g). Sweden and Finland were operating EMS-type vehicles before the formal introduction of EMS, and currently permit EMS vehicles up to 25.25 m in length and 60 tonnes gross mass [1]. The Netherlands, Germany and Spain have since also introduced 25.25 m EMS combinations (up to 60 tonnes in Netherlands and Spain, and up to 44 tonnes in Germany) [17, 40]. Denmark is trialling the introduction of 25.25 m EMS vehicle combinations up to 60 tonnes [17].

The Longer Semi-Trailer (LST) trial in the UK [7] is evidence of the UK's growing interest in HCVs. The 10-year trial was launched by the Department for Transport (DfT) in 2012 with up to 1,800 vehicles participating under special permits. The trial permits an additional 2.05 m to the length of the semi-trailer, allowing additional loads to be carried, and thereby reducing the number of truck trips and kilometres travelled. As at 31 Dec 2015, 1,747 vehicles were participating in the scheme.

Official data collected during the LST trial show that the use of the longer semi-trailers has saved 8.7–10.6 million vehicle kilometres, removed 1 in 19 vehicle trips, and exhibited 70% fewer personal injury accidents compared with standard articulated HGVs. In January 2017, the DfT announced that the trial would be extended by 5 years, and the number of participating vehicles increased to 2,800 [41].

The trend towards larger and heavier vehicles is also evident in conventional HGV operations. In the UK in the period from 1990 to 2015 [7], there was an 83% increase in the use of rigid trucks above 25 tonnes, while the use of rigid trucks below 17 tonnes decreased

by 76%. Similarly there was an increase of 69% in the use of articulated HGVs over 33 tonnes, while the use of those below 33 tonnes decreased by 86%.

1.4 Enabling technologies for High Capacity Vehicles

Long HCVs with multiple points of articulation exhibit limited manoeuvrability in terms of turning and reversing. They may also exhibit reduced yaw stability compared with rigid or singly-articulated vehicles, resulting in increased rearward amplification of lateral acceleration (sometimes leading to trailer rollover), increased transient off-tracking of trailers (overshoot of the rearmost trailer into adjacent lanes during an evasive manoeuvre), or jackknifing [42]. Addressing these issues is an important factor in enabling the uptake of HCVs.

To some extent, the manoeuvrability of HCVs can be improved through careful selection of passive vehicle design parameters such as trailer configuration, axle positions and hitch locations. The scope for improvement is very limited however, given constraints of practicality, payload type and geometry, and maintaining acceptable axle load distribution. Furthermore, changes to these parameters which favour manoeuvrability are often at the expense of high-speed safety performance characteristics [42–45].

Similarly, the high-speed safety performance of HCVs can be improved through the careful selection of passive vehicle parameters such as axle and hitch locations. Some changes such as increasing the wheelbase of trailers can improve high-speed stability while at the same time improving axle load distribution and also bridge loading and road wear impact. However, the effects of additional trailers can only be addressed to a limited degree through passive design choices. The specification of suspension components and tyres can also effect performance, particularly in improving rollover performance and reducing dynamic tyre forces, but there is generally very little scope here for improving manoeuvrability and high-speed dynamic performance.

There is hence a need for active intervention technologies to address the manoeuvrability and stability limitations of HCVs, given the limitations of passive design choices [25, 46, 19, 22]. Such systems include autonomous reversing, anti-jackknife control, active trailer steering and combined braking and steering control. These technologies are beneficial to improving the accessibility and safety of HGVs in general, but are particularly important in enabling wider use of HCVs on more of the road network.

Some of these technologies will now be described in more detail, highlighting the current state of development and some important limitations.

1.4.1 Autonomous reversing

The reversing of an articulated vehicle is an inherently unstable process, and so some level of control is needed to achieve a desired reversing task [47]. For simple articulated vehicles such as tractor semi-trailers, an experienced driver is able to perform this control task, manually providing control inputs to the front steering wheels of the tractor in response to the observed response of the trailer.

For vehicles with two or more points of articulation, the control task becomes exceedingly challenging for a driver to perform, given the increased number of unstable degrees of freedom and the non-linearity and unpredictability of the system. This has warranted a number of investigations into the development of suitable automatic controllers for this task. These controllers can be in the form of either reverse assist or fully autonomous reversing where the required steer inputs are calculated and input automatically. The current discussion will focus on autonomous reversing.

Some recent work on the topic includes a 'virtual tractor'-based controller [48], a 'unicycle' controller for short single-axle trailers [49], and the controller of [50] which is limited to certain axle and hitch configurations. These systems, together with the majority of other work on the topic, are predominantly theoretical in nature, difficult to tune in reality, and their validity has not been evaluated on full-scale HGVs [29].

More recently, Rimmer *et al.* [47, 29, 23] demonstrated practical controllers for the reversing of multiple trailer HGVs, and successfully demonstrated their feasibility in several full-scale tests. In [29], Rimmer and Cebon demonstrated a state feedback controller, and evaluated it on a full-scale tractor semi-trailer, B-double, and B-triple vehicle combinations. In [23], a preview-point controller was presented, which offered reduced steer input and swept path at the expense of path-following error. In comparison tests with a B-double the two controllers were able to achieve path-following of the rearmost axle to within 50 mm and 220 mm respectively [23].

All the controllers proposed in [48–50, 29, 23] assume passive trailers, using only the tractor steer angle as the control input. In terms of sensing, Rimmer's reversing controller [29, 23] requires measurements of steer angle, position and heading of one of the vehicle units (*i.e.* the tractor), and articulation angle. This can hence be viewed as tractor-based technology, with the exception of articulation angle sensing.

1.4.2 Anti-jackknife control

If either a truck or trailer experiences yaw instability, normally resulting from a loss of traction of one or more axles, then yaw instability of the entire vehicle combination will

result [51]. There are two ways in which this can occur, known as 'tractor jackknife' and 'trailer swing', depending on which axles lose their lateral grip. This often results in the tractor and trailer folding in towards each other, commonly known as 'jackknifing'. Jackknifing can also occur during reversing due to the inherent instability of articulated vehicles in reverse.

Various jackknife prevention systems have been proposed, with three primary control actuation methods: hitch actuation, braking, or steering (see Figure 1.1b). Hitch actuation can be in the form of articulation locking or articulation damping [51]. Jackknife control via braking can be achieved either by removing wheel lock using ABS[7] or EBD[8] to produce correcting yaw moments [25]. Alternatively, steering inputs from the tractor can be used to correct jackknifing [25].

Azad *et al.* [51] investigated the dynamics of jackknifing, and the role that articulation damping could have in prevention. Bouteldja *et al.* [30] developed a jackknife detection and prediction system for tractor semi-trailer combinations based solely on the observation of articulation angle. A state observer and an Extended Kalman Filter (EKF) were adopted to estimate the angle given measurements of steer angle, lateral velocity and yaw rate, and the system was assessed through simulation.

Chu *et al.* [52] highlighted that knowledge of the articulation angle is fundamental to any jackknife detection system. They identified a lack of inexpensive and practical articulation angle sensing technology, and so introduced a state observer to estimate it for a tractor semi-trailer using other sensors more likely to be available on commercial vehicles. The system required measurements of steer angle, lateral acceleration, yaw rate, and axle loads, most of which would be available from an Electronic Stability Control (ESC) system if fitted. Performance was shown to be good, except in instances of low tyre-road friction, introducing the additional need for friction estimation.

Chen *et al.* [25] developed a differential braking scheme to prevent jackknifing in a tractor semi-trailer using model reference adaptive control. The system relied solely upon the articulation angle for control feedback. The system was validated in computer simulations and scale model tests. Chiu *et al.* [53] conducted a study of jackknifing to determine the critical articulation angles for the onset of jackknifing, particularly in reversing manoeuvres.

It is clear that articulation angle is the most important variable of concern for jackknife detection and control, regardless of the actuation and control methods employed.

[7] Anti-lock Braking System
[8] Electronic Brakeforce Distribution

1.4.3 Combined braking and steering control

Anti-lock Braking Systems (ABS) serve two purposes: preventing the lock-up of wheels under severe braking and hence allowing a driver to maintain directional control, and reducing stopping distances by maintaining higher static rolling friction between the wheels and road surface [24]. For HGVs, the technology has been shown to be less effective in reducing stopping distances than in the case of passenger vehicles, as a result of the sluggish response time of the pneumatic braking systems on HGVs compared to the hydraulic braking systems on passenger vehicles [24].

'Slip-control' braking has been shown to significantly improve stopping distance compared to ABS through the use of fast-acting pneumatic valves and accurate tyre slip modulation. Successive investigations by Kienhöfer *et al.* [54, 55], Miller *et al.* [56, 46], and Henderson *et al.* [57, 27] resulted in the design of a novel fast-actuating pneumatic valve and sliding mode tyre slip controller. Recent field tests demonstrated a 17% reduction in stopping distance and 30% reduction in air consumption during braking on a low friction wet basalt-tile surface when the braking system was fitted to the trailer [58].

Morrison showed that although the slip-control system was very effective in reducing stopping distances, this was at the expense of directional stability compared with conventional ABS [24]. Morrison *et al.* [59, 60, 24, 31] extended the slip-control braking system to address scenarios of combined braking and steering, and introduced the 'attenuated slip demand' concept. This yielded both directional control and stopping distance performance superior to that of conventional ABS. Morrison's work on combined braking and steering control is important for the long term commercial prospects of slip control braking, but the attenuated slip controller introduces additional sensing requirements. Sensor signals required for attenuated slip demand include: forward speed, side slip, yaw rate and steer angle from the tractor, and importantly also requires articulation angle.

In [59, 60, 24, 31], articulation angle sensing was achieved with a trailer-based 'VSE' sensor (see Section 1.5) with a CAN-based communications network connecting the controllers in each vehicle unit. Although the fast acting valves are not yet standard hardware, it is reasonable to assume that the next generation of HGV braking hardware will adopt such faster-acting valves, and make use of slip-control and combined braking and steering technologies such as these. This leaves the significant problem of measuring articulation angle and transmitting it to the master braking controller in the tractor unit. Measuring articulation angle on the tractor unit would solve this problem.

1.4.4 Trailer steering

Off-tracking refers to the lateral deviation of the rear of a truck or trailer as it navigates a turn, or as it is subjected to low friction and/or cross slope. Off-tracking is illustrated in Figure 1.3a for a tractor semi-trailer. The magnitude of off-tracking can be measured in a number of ways. In this research it is taken to be the lateral deviation of the rear of the trailer relative to the path of the hitch point, measured laterally to the trailer axis.

'Tail swing' is another important manoeuvrability metric which pertains to the level of swing-out of the rear of a truck or trailer during a turn. Importantly, tail swing occurs on the outside of a turn, where the driver has no line of sight. This is usually the result of a short wheelbase and/or large rear overhang, a common feature of centre-axle trailers. Excessive tail swing can lead to collisions with roadside furniture and parked vehicles, stationary vehicles in adjacent lanes, or with cyclists or pedestrians.

Conventional trailer axles are unsteered. This limits the overall manoeuvrability of the vehicle combination, and also results in tyre scrub for trailers with multiple unsteered axles. A number of trailer steering technologies exist which aim to remedy one or both of these issues, and these can be generally classified into passive and active systems.

Passive steering systems, in order of complexity and cost, include: 'self-steering', 'command steer', and 'pivotal bogie' systems [61]. Self-steering offers the least benefit in terms of manoeuvrability and tyre scrub. Command steer and pivotal bogie systems can significantly improve manoeuvrability (through reduced cut-in) and tyre scrub, but tend to exhibit excessive tail swing behaviour in transient manoeuvres. Passive steering systems have also been shown to degrade stability at high speeds, and usually require locking in place above a threshold speed [28].

By comparison, active steering systems can approach optimal manoeuvrability performance without unfavourable transient effects, and have also been shown to improve high-speed safety performance [32, 28, 62]. An early active steering concept was proposed by Hata *et al.* [63] to reduce the off-tracking of rigid trucks without increasing tail swing. This concept was extended by Notsu *et al.* [64] for a steered semi-trailer, in which the rear of the trailer actively followed the path of the front of the tractor. These are early examples of the 'path-following' trailer steering method.

More recently, Jujnovich and Cebon [21] developed an active path-following semi-trailer steering system in which the rear of the trailer follows the path of the fifth wheel for all speeds and paths. This is illustrated in Figure 1.3b. Their strategy for achieving this was to compare heading angles at the front and rear of the trailer using a non-linear control strategy. A 'model-matching technique' requiring a reference trailer model was used to address problems associated with steering angle limits at low speeds. The controller required

knowledge of vehicle parameters from both tractor and trailer, as well as measurements of speed, articulation angle and trailer yaw rate. The performance of the path-following concept compared with a conventional trailer and command steer is shown in Figure 1.4, highlighting its superior cut-in and tail swing behaviour.

An alternative strategy was proposed by Cheng [32, 20], using a 'virtual driver' at the rear of the trailer and a simpler proportional–integral–derivative controller to determine off-tracking at low speeds. The system was also applicable to high speeds, minimising a cost function with respect to high-speed stability instead of off-tracking. The controller required knowledge of various tractor and trailer parameters, as well as measurements of tractor speed and steer angle, trailer yaw rate and articulation angle. The controller assumed zero wheel sideslip at low-speeds.

By basing the control on off-tracking error instead of heading angles, Cheng's approach avoided possible steady-state tracking errors that are possible in Jujnovich's approach. Further, Cheng's controller does not require the reference trailer approach to handle steer saturation. A prototype trailer incorporating the path-following system is shown in Figure 1.5, developed by the Cambridge Vehicle Dynamics Consortium (CVDC). The trailer currently uses Jujnovich's control strategy, as it is currently better suited to low speeds and simpler to implement, but may incorporate Cheng's controller in future. Cheng's work was subsequently extended by Roebuck *et al.* [22] for multiple trailers.

Although Cheng's controller improved on certain aspects of Jujnovich's, it has two particular shortcomings relevant to the current work. Firstly, like Jujnovich's controller, it requires measurements from the tractor, and also knowledge of tractor parameters for the vehicle model. Secondly and importantly, the assumption of zero slip limits the application of the system to on-highway and high friction conditions. This is a reasonable assumption for many applications, but is limiting for icy conditions on the highway and for off-highway applications such as logging, livestock and dairy collections from farms, and military convoys. Off-highway applications are prone to low friction, cambered and inclined roads, and tight corners, giving rise to potentially large wheel slip.

Miao and Cebon [65] demonstrated how the performance of Cheng's path-following control can be severely degraded in conditions of low friction, camber and grade. Simulating a path-following tractor semi-trailer combination traversing a 450° roundabout manoeuvre, they showed a 0.4 m increase in off-tracking error when 5° road camber and 5° grade were introduced (Figure 1.6a). A manoeuvre with 5° camber and a co-efficient of friction of 0.1 resulted in a 0.6 m increase in off-tracking error (Figure 1.6b).

Miao [3] showed how Cheng's controller could be reformulated to take an independent measurement of off-tracking, *i.e.* relaxing the no-slip assumption. This paved the way for the investigation of alternative off-tracking estimation methods.

1.4.5 Barriers to adoption

The technologies just described are promising enablers of HCV uptake, addressing important limitations of these vehicles. However, some important practical challenges associated with these technologies have also been highlighted, particularly their sensing requirements.

It is clear that articulation angle sensing is a fundamental requirement for autonomous reversing, anti-jackknife, and combined braking and steering technologies. Existing articulation angle sensors inhibit the feasibility of these technologies, in that they are either trailer-based or require some knowledge of trailer parameters.

Similarly, trailer steering is a highly effective technology for improving the manoeuvrability of longer HCVs, especially in challenging environments. The most promising such technology is the CVDC's path-following concept [32, 20], which has been shown to achieve an optimal balance between cut-in and tail swing, while reducing tyre scrub. However, the system assumes knowledge of tractor parameters, and Miao [65] highlighted its sensitivity to reduced friction levels, deeming it potentially ineffective in icy or off-highway conditions.

The wheel slip dependence originates from the manner in which off-tracking is estimated. Finding a solution for trailer off-tracking estimation which is independent of friction levels and which requires no knowledge of tractor parameters would go a long way towards improving the performance and commercial feasibility of effective trailer steering technology.

In the following sections, articulation angle and trailer off-tracking measurement will be discussed in more detail. Existing technology and the latest research are described, highlighting the shortcomings and scope for further work.

1.5 Articulation angle sensing

At the start of this discussion, it is useful to establish the precision which an articulation angle sensor would require for application to HGVs. The reversing controller of Rimmer *et al.* [47, 29, 23] tries to ensure that the rear of the trailer follows a user-defined path as accurately as possible, with a target 'path following error' of zero. For reasonable obstacle avoidance and control stability, it is estimated that actual path-following errors should be no more than 0.3 m (peak-to-peak). Assuming a zero-mean sinusoidal error signal, this translates into a maximum error of 0.15 m and an RMS error of 0.10 m, in line with previous

work on path-following steering control [32, 21]. Articulation angle sensing is used by the controller to calculate the current position of the rear of the trailer relative to the tractor using trigonometry, from which the current path-following error is calculated. Therefore, given target the path-following error statistics, the required articulation angle error statistics may be calculated. For a UK Longer Semi-Trailer with a distance of approximately 14 m from kingpin to trailer rear, this translates into articulation angle errors of $\sin^{-1}\left(\frac{0.15}{14}\right) = 0.6^\circ$ maximum, and $\sin^{-1}\left(\frac{0.10}{14}\right) = 0.4^\circ$ RMS. RMS is a useful accuracy metric as it incorporates both mean and noise error components, and will be used throughout this thesis.

1.5.1 Commercial and experimental sensors

Various articulation sensors exist either commercially or for research and development work. Examples include: a kingpin sensor commercially available from Vehicle Systems Engineering B.V. Netherlands (the 'VSE sensor') [66], a prototype Orbisense™ magnetic sensor by AB electronic Ltd [32], and a custom string potentiometer solution [28] (see Figure 1.7). TRIDEC (Netherlands) use a custom articulation angle sensor as part of their active trailer steering system [67].

These are all trailer-based sensors, and require non-standard communication links with the tractor. Furthermore, the VSE and Orbisense sensors require significant modifications to the trailer kingpin and the string potentiometer requires a physical connection between tractor and semi-trailer. The Orbisense™ sensor was never commercialised, for reasons including insufficient resolution for active-trailer steering applications [28]. With the exception of the Orbisense™ sensor, these sensors are examples of 'contact-type' sensors. For semi-trailers the fifth wheel is subjected to high static and dynamics loads and is a dirt and grease-prone environment, with potentially negative effects on the longevity of these sensors. For this application a good case is made for a non-contact sensing solution.

A small number of patents exist for articulation measurement for passenger vehicles with towed trailers [68, 69]. These systems include the use of multiple sensors, some of which include cameras, but the precision of these systems is unknown and, in cases where cameras are used, details of the image processing methods are not clear. The system proposed in [68] requires an extensive manual calibration procedure.

1.5.2 State observers

A state observer approach has been proposed in various studies, in which articulation angle is estimated from other sensor measurements coupled with a full motion model of the vehicle.

A drawback to this approach is the need for an accurate vehicle model, necessitating the assumption of numerous geometric and dynamic parameters.

Bouteldja *et al.* [30] developed a jackknife detection system for tractor semi-trailer combinations based solely on the observation of articulation angle. They adopted a state observer and an EKF for articulation estimation and assessed the system through simulation. Although results were not given for the precision of articulation angle measurements, the maximum error in estimated tractor yaw angle using the state observer approach was in the region of 8°. The system requires sensing of steer angle and tractor position.

The articulation angle observer of Chu *et al.* [52] (see Section 1.4.2) exhibited maximum errors of 0.3° for small angles (up to 3°), 1.6° in a low friction test up to 10° (with friction estimation enabled), and 10° in a low friction test up to 20° (with no friction estimation). The observer proposed by Ehlgen *et al.* [70] achieved a maximum error of 5.4° in vehicle tests with articulation angles up to 48°, requiring only a steer angle sensor input.

1.5.3 Camera-based sensors

The use of cameras to measure articulation angle has been proposed in the work of Schikora *et al.* [71], Caup *et al.* [72], Harris [73] and Fuchs [74, 75]. The favoured set-up consists of a camera, or cameras, mounted to the rear of the towing vehicle, observing the front of the trailer, and inferring articulation angle through suitable image processing methods.

Schikora *et al.* [71] proposed two vision-based methods for the measurement of articulation angle: one for semi-trailers and one for drawbar trailers. For semi-trailers they proposed using an encoder plate attached to the underside of the semi-trailer, which was viewed using a single upwards-facing camera beneath, shown in Figure 1.8a. The rigid drawbar trailer system requires specially-placed infrared diodes to the front trailer face, assuming the face is planar (if not, a planar mounting surface would be required). These diodes were viewed through a rear-facing infrared-filtered camera fixed to the rear of the hauling vehicle, as shown in Figure 1.8b. The encoder plate system has a claimed accuracy of around 0.3° for a full 180° range of angles, and a computation time of below 0.15 s, but details for the infrared system are not clear. Trailer modifications are required in each case.

Caup *et al.* [72] employed a single rear-facing camera to measure the articulation angle between a car and a simple drawbar trailer, with assumptions on the location of the hitch and the maximum length of the drawbar (see Figure 1.8c). The system requires no modifications to the trailer, but is applicable only to drawbar trailers. They utilised a template-matching method, trained to optimise parameter sets, and an alpha-beta filter to smooth measurements. In vehicle tests, average errors of 1.69–2.19° and maximum errors of 5.48–7.55° were

demonstrated for articulation angles up to 50°. The average time per computation was 0.15 s, making it real-time feasible.

In 2013, Harris [73] investigated a camera-based articulation angle sensing concept specifically for tractor semi-trailers. The concept consisted of a camera (or cameras) mounted to the rear of the tractor cab, facing rearward towards the front of the semi-trailer, and using image processing algorithms to extract the trailer angle from the image sequences. Harris' work is particularly relevant to the current research, and so is discussed in some detail here.

Harris investigated two image processing methods to accompany the rear-facing cab camera: 'homography decomposition' (using one camera) and stereo vision (using two cameras). The first method assumed a planar trailer front. A datum image of the trailer front at zero articulation was required, and subsequent images of the trailer were compared to this datum to determine the planar homography between the two trailer orientations. Feature matching was used to compare images, using SIFT[9] features and RANSAC[10] [76] (RANSAC is a model-fitting algorithm which is robust to the presence of many outliers). An example output of the matching process is shown in Figure 1.8d. The articulation angle was extracted from the homography using single value decomposition. A Kalman Filter was added to smooth the resulting noisy measurements.

The method was evaluated using image sequences generated by a 3-D CAD[11] model of a tractor semi-trailer combination. A resolution of 640×480 was used in the simulated images. Maximum errors of 8.4° were observed in tests up to 90° articulation angle. The largest errors originated from the transition from trailer front to trailer side in the camera view at large articulation angles. Errors of over 80° were observed without the Kalman Filter. The computation time was approximately 5 s per frame or 0.2 frames per second ('fps').

The stereo vision approach utilised a stereo camera pair and a block-matching image processing technique [77] to generate a disparity (depth) map of the trailer. The resulting 3-D point cloud was projected onto the horizontal plane, and a RANSAC scheme was used to extract the rectangular trailer outline and infer its angle (see Figure 1.8e). Although stereo vision is inherently three-dimensional, the method used to extract the trailer outline to determine the articulation angle assumes a box-type trailer with a 2-dimensional (planar) front. A Kalman Filter was again adopted to smooth the measurement output.

The stereo vision method was assessed using the same CAD simulation environment, for articulation angles up to 60°. Maximum errors of 1.2° were observed, up to 7° without the filter. The system was also implemented on a full-scale tractor semi-trailer, and slalom and roundabout manoeuvres were conducted with articulation angles of up to 60°. Maximum

[9] SIFT = Scale Invariant Feature Transform
[10] RANSAC = RANdom SAmple Consensus
[11] CAD = Computer-Aided Design/Drawing

errors of up to 18° were observed, or 26° without the Kalman Filter. Computation time was highly impractical at approximately 70 seconds per frame, but it is possible that the algorithm could be refined to improve computational efficiency.

Harris briefly introduced a third method, 'template-matching'. Assuming again a planar trailer, and a datum image of the trailer at 0° articulation angle, Harris showed that the datum image could be 'warped' to generate a database of images of the trailer at any angle, stored at 0.1° increments. Real-time images of the trailer could then be compared with the database to find the best match and hence best articulation estimate. 2-dimensional normalised cross-correlation was used to compare images. A very limited evaluation of the method was carried out, suggesting much improved accuracy and computation compared to the first two methods.

In 2014, Fuchs *et al.* [74] presented a camera-based articulation angle measurement concept for truck and full-trailers. The concept relied on three defined markers arranged in a triangular pattern fixed to the trailer front, assuming that the front of the trailer is planar. The system also required precise knowledge of vehicle geometry and camera location. A conceptual implementation of the system on a 1:16 scale vehicle is shown in Figure 1.8f.

The concept was tested in an idealised simulation environment (not on the scale truck shown in Figure 1.8f), using an image resolution of 640×480, for articulation angles up to approximately 50°. With perfect knowledge of camera location, maximum errors of 0.39° and 0.53° for dolly and trailer angles were obtained respectively. However, when a 6% error in camera location was incorporated (equivalent to 160 mm for a standard UK trailer), maximum errors rose sharply to 3.8° and 3.9° respectively, and increased linearly with additional camera location error.

In [75], Fuchs *et al.* added a Kalman Filter to their system, and again evaluated the system in an idealised simulation environment. Articulation angles of up to 15° and 30° for the dolly and trailer were investigated respectively. Maximum errors of 0.6° were obtained for the new system, versus 1.5° using the previous system. The Kalman Filter required the first 75 frames for initialisation.

In related work, in [78] a stereo vision SLAM[12] algorithm was proposed and used to measure the rotation angle between the shovel house and shovel car-body of a mining rope shovel. The system performed well and achieved maximum errors of under 1° over a rotation range of $\pm 180°$, though errors of up to 5° were observed before loop closure was reached.

In work investigating the navigation of small articulated robotic vehicles, Larsson *et al.* [79] used a rotating laser on the prime mover to detect reflective tape of known location affixed to the trailer. The measurements were coupled to an Extended Kalman Filter (EKF).

[12] SLAM = Simultaneous Localisation and Mapping (described in Section 1.7.2)

This could in theory yield acceptable accuracy (actual figures are not given), but is limited because it requires the existence of known markers on the trailer.

1.5.4 Performance summary

A performance summary of the most relevant articulation measurement concepts described above is given in Table 1.1. The methods of validation of these concepts can be separated into either simulation or full-scale vehicle tests, and so the performance results (maximum errors obtained, ε_{max}) should be viewed in this light. Maximum errors as a function of maximum articulation angles assessed during validation have been plotted in Figures 1.9a and 1.9b for simulation and vehicle test results respectively. (Shaded areas on the plot are a visual aid to help group data points from common sources.) Schikora's systems [71] have been omitted: the encoder system requires substantial trailer modifications, and accuracy information for the diode system is not clear.

1.6 Trailer off-tracking sensing

Miao [3] showed how Cheng's controller [32, 20] could be reformulated to take an independent external off-tracking measurement, and subsequently proposed a camera-based method for independent off-tracking estimation. To the author's knowledge, Miao's work is the only directly relevant work on trailer off-tracking estimation, and so is the focus of this section.

Target path-following errors were presented in Section 1.5 for the case of articulation angle sensing. Comparable targets for path-following accuracy are proposed here, as these relate directly to trailer off-tracking measurement. Using the UK Longer Semi-Trailer as a reference, maximum and RMS trailer off-tracking measurement errors should be 0.15 m and 0.10 m respectively for adequate control stability and obstacle avoidance.

1.6.1 Miao's ground-watching navigation system

To address the shortcomings in Cheng's controller, Miao [3] proposed a 'ground-watching navigation system' to measure off-tracking independently of wheel slip conditions. The work was primarily aimed at off-highway applications such as livestock and dairy collections, military convoys and logging.

The ground-watching concept was demonstrated for a tractor semi-trailer combination using two downwards-facing cameras mounted beneath the semi-trailer, one near the fifth wheel and one near the trailer follow-point. One of the cameras is shown in Figure 1.10.

Miao used FAST[13] features and SURF[14] descriptors to detect features in each image. Feature matching between front and rear camera images was achieved using the FLANN[15] algorithm. A RANSAC scheme was adopted to calculate the planar homography from the feature matches, and single value decomposition was used to determine planar rotation and translation from the homography.

For small off-tracking errors both cameras were used. Images from the front camera were stored in a data buffer, and compared to rear camera images when one trailer length had passed some time later. Calculating the planar homography between these images gave a direct measurement of off-tracking. Miao referred to this as the 'Dual Camera Ground Watching Navigation System' or DC-GWNS. An outline of the DC-GWNS is shown in Figure 1.11a.

When off-tracking errors were larger than the camera field of view such that front and rear images did not overlap (0.28 m lateral off-tracking error in this case), only the front camera was used. Planar homographies were obtained between sequential images, giving measurements of side-slip and lateral velocity at the fifth wheel. These data, together with trailer yaw rate and articulation angle measurements, were used with a yaw-plane vehicle model to estimate off-tracking at the end of the trailer. Miao referred to this as the 'Single Camera Ground Watching Navigation System' or SC-GWNS. An outline of the system is shown in Figure 1.11b.

Performance of the two systems was demonstrated in vehicle tests on dry tarmac with a tractor semi-trailer. Off-tracking measurement errors of 0.05 m were obtained at 10 Hz in open-loop tests (*i.e.* with non-steered trailer axles and no control). In closed-loop tests, using measurements from the GWNS as inputs to Miao's modified path-following controller, path-following errors of less than 0.1 m were obtained.

Although the system addressed the issue of wheel slip dependence, it assumed an unchanging and planar road surface with static features. In conditions where wheel slip is likely (*i.e.* where there is mud, standing water, ice, snow), it is also likely that the road surface is non-planar and varies with the passing of the vehicle, and so may be unsuitable for the ground-watching concept.

The sensitivity of Miao's system to these conditions was not explored, as tests were conducted on dry tarmac. Standing water would potentially invalidate the feature-matching process (due to the reflection of the trailer undercarriage). Also, a soft muddy road surface will be subject to displacement and settling as the vehicle passes over it, meaning that the

[13] FAST = Features from Accelerated Segment Test
[14] SURF = Speeded Up Robust Features
[15] FLANN = Fast Library for Approximate Nearest Neighbours

surface seen by the front camera may differ from that of the rear camera which would invalidate one of the fundamental assumptions.

From a practicality point of view, cameras mounted beneath a trailer in these conditions would also be very prone to dirt and water splash which could significantly affect the performance of feature extraction and matching. Although this risk exists for most outdoor camera-based measurement applications, the under-trailer configuration has a far greater exposure than cameras mounted elsewhere on the vehicle.

These unfavourable road conditions and their implications for camera-based off-tracking measurement are an important consideration in current and potential off-highway applications for trailer steering, and so should be taken into account in further developments of this work.

1.6.2 Other related work

Other applications of path-following in high slip environments include agricultural vehicles and planetary exploration. Cariou *et al.* [80] developed a system to guide a towed agricultural implement (*i.e.* a trailer) along a predefined path, controlled with steer inputs from the towing vehicle. The system yielded off-tracking errors of approximately 1.7% of distance travelled in scale model tests with low friction.

The system utilised an articulation angle sensor, a kinematic vehicle model, and high precision RTK-GPS (Real Time Kinematic GPS). This is a method of GPS positioning wherein an additional GPS base station is used to make corrections to normal GPS measurements, increasing accuracy significantly. This is not practical for road vehicle applications as a line of sight with the base station must be maintained to receive corrections.

Helmick *et al.* [81] proposed a visual odometry-based system for the global path-following of a Mars rover in a high-slip environment. Stereo visual odometry data was merged with data from an inertial measurement unit using a Kalman Filter. This was complemented with kinematic measurements based on wheel speed and steer angles. The system was able to achieve 2.5% path-following accuracy in manoeuvres on loose sand and cross-slopes of up to 15°.

1.7 Pose and ego-motion estimation

Fundamentally, the two tasks of articulation angle and trailer off-tracking measurement are problems of pose estimation and ego-motion estimation. Pose estimation generally refers to determining the position and orientation of an object which is external to the

observer/sensor. Ego-motion estimation refers to the measurement of an observer's own position and orientation.

Pose and ego-motion estimation are important in many applications, including the navigation of mobile robots and Unmanned Aerial Vehicles (UAVs), human pose estimation and gesture recognition, and multiple aspects of autonomous and semi-autonomous passenger vehicles such as navigation, lane keeping, obstacle avoidance and vehicle dynamics control.

The following discussion provides an overview of existing vehicle-based sensing technologies which are currently used, or could be used, for pose and ego-motion estimation tasks. This is followed by an overview of some important pose and ego-motion data processing methods used in vehicle and related applications.

1.7.1 Vehicle-based sensing technologies

Modern passenger vehicles require a number of sensing tasks in order to interact safely with the environment and with other vehicles and road users. This is particularly true of autonomous and semi-autonomous vehicles. Typical sensing tasks include obstacle detection, navigation, pose estimation, and ego-motion estimation. These tasks are often inter-related, and the sensing technologies for each can be complimentary and interchangeable.

First, a brief mention of GPS (Global Positioning System), as this is commonly used for vehicle and robot localisation tasks and is widely available. GPS is able to provide global positioning information, accurate to between a few metres and a few centimetres, and so does not suffer from drift problems. Pose and ego-motion measurement with GPS is possible through the use of multiple GPS receivers, and this is commonly used for vehicle testing, but in order to achieve centimetre-level accuracy high precision differential GPS would be required, and this is both costly and impractical for on-road applications.

Perhaps the simplest direct method of ego-motion estimation and navigation is wheel odometry [82]. Wheel encoders are used to determine the rotation of individual wheels on a vehicle or mobile robot, and from these measurements the motion and position of the vehicle can be calculated. This is a very low cost solution, but drawbacks include unbounded errors due to integration and the assumption of no wheel slip. The use of wheel odometry is nonetheless very common in mobile robots, and is often used to compliment other sensors.

Ultrasonic sensors are commonplace on passenger vehicles, commonly used as parking sensors. Ultrasonic sensors rely on the transmission and reflection of short wavelength sound waves (with wavelengths of a few centimetres) to provide non-directional depth measurements (to a parked vehicle for example). These sensors are low cost, and the short wavelength provides good accuracy at low range. However, their maximum range is limited to a few metres.

Ultrasonic sensors have also been used for cyclist and pedestrian detection applications. Existing commercial systems include the 'Autowatch' [83] and 'Backwatch' [84] systems. Recent research developments include work by Jia [85, 86] in which an array of ultrasonic sensors was used to detect cyclists next to an HGV and also to predict their future motion.

Radar (RAdio Detection And Ranging) is increasingly common on passenger vehicles. Common applications include adaptive cruise control, collision detection and avoidance, and lane change assistance. Radar utilises the transmission and reflection of long wavelength radio waves, giving it good range (up to a few hundred meters). Radar is more costly than ultrasonic sensors and is sensitive to the porosity of the target surface, passing though porous surfaces such as plastic bumpers instead of reflecting.

Early vehicle-based radar technology was limited to 77 GHz Long Range Radar (LRR), which has a good range (10–250 m) but relatively poor resolution (0.5 m) [87]. LRR is hence primarily used for forward facing radar for adaptive cruise control. Later, 24 GHz Short Range Radar (SRR) was introduced with a range of 0.15–30 m and a resolution of 0.1 m [87], enabling its use for blind spot detection, lane change assist, and some parking assistance applications. Combinations of ultrasonic sensors and long- and short-range radar are often used for modern passenger vehicle applications, such as on the semi-autonomous Tesla Model S. An illustration of the typical sensors and their purposes on a modern passenger vehicle is given in Figure 1.12.

Lidar (LIght Detection And Ranging) has been used in many applications including robotics and cartography, primarily for high-density three-dimensional mapping [88]. Lidar operates in a similar manner to radar but utilises the transmission and reflection of light waves of varying wavelengths to detect distances. In modern lidar systems, it is common for multiple lasers to be aligned vertically and rotated at high speed to give a reasonable vertical field of view and up to 360° horizontal field of view.

Lidar began as a mapping and ranging technology for geographers, but modern lidar has found significant use in mobile robotics and autonomous vehicles. The use of lidar for autonomous vehicles has been showcased by prominent autonomous vehicles including Google's self-driving car and several DARPA Grand Challenge entrants [89]. The Velodyne HDL-64E S2 used on Google's car has a $360^\circ \times 26.8^\circ$ field of view, 0.09° angular resolution, and 2 cm distance accuracy [90]. These specifications make it attractive for one-off prototypes, but its $80,000 price (as of 2013) [91] limits its commercial feasibility for most vehicle applications. The Google car and Velodyne sensor are shown in Figure 1.13.

Ultrasound, radar and lidar are all examples of 'active' sensing technologies. They emit a signal under their own power and use information from the reflected signal to infer distance. Active sensors are attractive in that they can measure distances and motion directly,

utilise very little data, and require minimal processing power. The lack of data allows for simple processing, but also limits the amount of useful information which can be extracted. Additionally, ultrasonic and lidar sensors are sensitive to weather conditions. Another important consideration with active sensors is that they suffer from interference from other active sensors of the same type. Given the increasing use of vehicle-based sensing for intelligent vehicles, an abundance of active sensors would be problematic [92, 93].

Vision-based sensing

Vision-based sensing refers to the use of cameras and appropriate image processing techniques to obtain information from a scene. Vision-based sensors are passive and only process light received which has been reflected from objects in the environment. With appropriate processing techniques, images can be used to infer both spatial information such as pose and ego-motion, but also qualitative information such as identifying pedestrians or traffic signals.

Images are extremely rich in information, and their ability to provide both spatial and qualitative information has resulted in the widespread uptake of vision-based technologies in the automotive sector and more broadly. Compared to active sensing technology, vision technology requires significantly more processing to extract the data of interest which can be complex and computationally demanding. However, significant developments in image processing algorithms and computer processing power have made modern vision-based technology feasible and affordable for a wide range of real-time applications [94].

It is widely accepted that drivers rely predominantly on visual information [95], and so the use of artificial vision systems for vehicles is intuitive. In a vehicle context, cameras are simple to implement, versatile and relatively inexpensive, and not prone to sensor-to-sensor interference, making them commercially attractive [96–98]. Additionally, existing road infrastructure provides many visual cues such as lane markings and road signs [92], and vision-based systems are well suited to differentiating between objects such as pedestrians and cyclists [93], further increasing its attractiveness to the automotive sector.

Extensive overviews of the developments and applications of computer vision to road vehicles are given by Bertozzi *et al.* [92] and by Dickmanns [94].

Vision-based state estimation has been explored in both UAVs and road vehicles, including roll angle estimation for two-wheeled vehicles [99, 98]; attitude and motion estimation for UAVs [100–102]; and side-slip estimation for ground vehicles by [103, 104]. Vision-based obstacle detection for both on and off-road vehicles has been studied extensively [105–110].

Vision-based sensors are often used in conjunction with other sensors such as radar and ultrasound, so that they may complement each other, and provide additional factors of safety. When all sensors are available and operational, complementary sensors can be

combined through 'sensor fusion' [93]. This is often achieved though the use of Kalman Filter techniques. Using cameras in a sensor fusion approach is common in autonomous and semi-autonomous vehicle applications including Google's self-driving car and the Tesla Models S, X and 3. Modern cyclist detection technologies also utilise a combination of cameras and radar, such as the system of Volvo Group [111] and Fusion Processing's CycleEye® [112]. Image data from the camera/s are used for detection, while radar data gives accurate depth measurements. The combination of sensors in such a fashion can be costly, and so much research has been conducted into solving vehicle-based sensing problems using vision technology alone.

The number of image processing methods developed for vision-based sensing, particularly for pose and ego-motion estimation, is significant. Three methods are particularly relevant, and have been used extensively in comparable tasks previously: Simultaneous Localisation and Mapping (SLAM), visual odometry, and template-matching methods. These three methods will be covered in more detail in the following sections.

1.7.2 SLAM

Simultaneous Localisation And Mapping or 'SLAM' is the problem of determining the position of a moving agent (vehicle, robot, human observer) while at the same time constructing a map of its environment. The solution to this problem has far-reaching applications to road vehicles, mobile robotics, unmanned aerial vehicles, and augmented reality. In general, SLAM involves the acquisition of sensor data, detecting features or 'landmarks' in these data, finding associations in these features between frames (either ordered or unordered frames), and using probabilistic models to determine the optimal solution to both localisation and mapping problems which minimises errors.

The first successful SLAM implementation utilised an EKF and has come to be known as EKF-SLAM [113], though many variations and implementations of this algorithm have since been developed. More recent developments include FastSLAM [114] and FastSLAM 2.0 [115], which uses a particle filter instead of a Kalman Filter, making it more effective for non-linear motion models and more computationally efficient. Other variants include Graph-based SLAM [116] and Topological SLAM [117].

The sensing input for SLAM can come from a variety of sources, as long as the sensor can detect and locate features or landmarks. Using inputs from lidar or cameras is common, but radar and ultrasonic sensors have also been used [118]. More recently, work on visual SLAM (or 'vSLAM') has dominated the literature, owing to the abundance and low cost of cameras, and significant improvement in vision processing algorithms and processing power.

Early visual SLAM utilised stereo cameras (stereo SLAM), but more recent developments have made SLAM possible with a single camera (mono SLAM) [119]. Mono SLAM gives position measurements up to an arbitrary scale factor[16] unless provided with additional information. Visual SLAM can be either feature-based, detecting and tracking individual image features, or dense, tracking pixel-level variations.

Important vSLAM algorithms include ORB-SLAM [120], a sparse mono SLAM using ORB[17] features, which has since been superseded by ORB-SLAM2 [121] which can take mono, stereo, or RGB-D inputs. (RGB-D sensors will be described shortly.) LSD-SLAM[18] [122] is a dense mono SLAM algorithm which gives rise to higher accuracy in low texture environments and dense scene reconstruction. Both ORB-SLAM and LSD-SLAM algorithms have been made open-source by their authors.

Parallel Tracking and Mapping (PTAM) [123] is a novel and highly efficient implementation of mono SLAM, in which the localisation and mapping processing tasks are formulated in parallel. The PTAM algorithm was developed for 'augmented reality' applications, and so is best suited for use within a localised area as opposed to for exploration. The parallel formulation results in excellent computational efficiency, while allowing for computationally expensive pose and map refinement to take place when other mapping tasks are not needed. 'Bundle adjustment' [124] is used for this task. PTAM is hence highly accurate, but can only provide position estimates up to an arbitrary scale factor.

The parallel SLAM formulation introduced by PTAM has since been adopted in a number of other SLAM implementations. Notably, S-SLAM [125] is a stereo implementation of PTAM which removes the scale ambiguity of PTAM, but which is less computationally efficient. The source code for both PTAM and S-PTAM algorithms is open-source [126, 127].

For navigation applications, recent developments have shown that the mapping task of SLAM can be almost entirely replaced through the use of pre-existing maps. The detail and availability of street level maps (such as Google Street View) and public photographs have allowed for surprisingly accurate pose and location estimates to be made without the need for the traditional mapping task within SLAM. A notable example of this is PoseNet [128], which can estimate location and pose from a single monocular image using Convolutional Neural Networks (CNNs).

[16] 3-D measurements which are accurate 'up to a scale factor', are measurements which are known only in a relative sense and not an absolute sense. For example, a single camera SLAM system is able to generate a scaled 3-D map of feature points from multiple views of a scene, but it is not known to the camera whether the 3-D distances between points are in centimetres, meters or kilometres. To determine scale, the change in camera pose between frames must be known, which can be obtained with stereo cameras or from external measurements such as wheel odometry.

[17] ORB = Oriented FAST and rotated BRIEF

[18] LSD = Large-Scale Direct Monocular

In any discussion of SLAM, it is important to mention the concept of 'loop closure'. This stems from a fundamental navigation problem of SLAM, in that errors in pose and position are unbounded with time due to the accumulative nature of the motion estimates. In a typical loop closure mechanism, when a mobile robot returns to a previously explored area and previously detected landmarks are recognised, current pose and location estimates are mapped with previous data from that location in order to refine current and historical estimates. Various approaches have been proposed for this, with varying levels of accuracy and practicality [129], though bundle adjustment is commonly used for this task.

Finally, RGB-D sensors have attracted some interest in the SLAM community. An RGB-D sensor is a combination of a colour camera with a depth sensor. A low-cost commercial example of this is Microsoft's Kinect™ sensor which uses an infrared emitter and sensor and a low resolution colour camera. Khoshelham and Elberink [130] determined the error and resolution of the Kinect™ sensor to be quite low at 4 cm and 7 cm respectively at the maximum range of 5 m.

1.7.3 Visual odometry

Visual odometry is effectively a subset of vSLAM, in which the focus is primarily on the localisation task. In addition, where the focus of SLAM is primarily on global localisation and optimisation by using unordered frames (*i.e.* not necessarily sequential), visual odometry is incremental, utilising sequential (ordered) frames, and optimises motion estimates only over the most recent frame, or a few of the most recent frames.

This is an attractive alternative to SLAM if pose accuracy is of primary importance, and especially if only short term motion is of interest. Should global position estimates be required, unbounded drift would result in large errors over time, and would need to be compensated with additional sensors such as GPS. Visual odometry algorithms employ various unique mechanisms to minimise drift, with varying degrees of effectiveness.

For mobile robots and vehicles, visual odometry essentially serves the same task as wheel odometry, but with better accuracy and without any dependence on wheel slip conditions [82]. As with vSLAM, visual odometry can be either mono or stereo, with mono implementations being typically less accurate and position estimates only true up to an arbitrary scale factor.

Fundamentally, the visual odometry procedure requires calculating the transformation in camera pose between two sequential images (in terms of translation and rotation). This is typically done by identifying and matching features between images, calculating the relative pose of the camera through triangulation and a perspective camera model, and minimising reprojection errors over all detected features using a suitable optimisation scheme [82].

Scaramuzza and Fraundorfer [82] give an extensive review of the background, literature and methods of visual odometry for the interested reader. The number of existing visual odometry algorithms is vast, with varying degrees of accuracy, computational efficiency and practicality. An extensive list of algorithms can be found in [131]. These include SOFT [132] (stereo), VISO2 [133] (mono and stereo), and SVO (mono and stereo) [134].

Visual odometry as it has just been described is a feature-based algorithm. Dense visual odometry is usually more commonly referred to as 'optical flow', and is more relevant to mapping applications, or where particular structural features of the environment are of interest (as in the case of pedestrian detection for example).

As with vSLAM and other visual methods, visual odometry is often fused with wheel encoders, GPS or Inertial Measurement Units (IMUs) to improve accuracy or reduce drift. However, as before there has been much focus on developing and improving vision-only methods so as to reduce reliance on expensive GPS and IMUs, or for applications where GPS cannot be guaranteed.

1.7.4 Template-based pose estimation

Template-based methods have been used in a number of applications to estimate the pose of an observed object, including vehicle detection and human pose and gesture recognition. Some prior knowledge of the observed object is required, such as a full 3-dimensional model (*e.g.* [135, 136]), or a number of template images of the object from different view points for which the pose is known (*e.g.* [137]). The task of determining the pose of the object becomes a task of recognition or 'matching', and hence techniques for this are often known as 'template-matching'.

The registration process may be carried out through direct pixel-based comparison between observed and template images, or through feature-based methods (see [138] for an in-depth survey of existing methods). Pixel-based registration techniques include cross-correlation [139], mutual information [140] and histogram matching [141]. Feature-based methods make use of distinct features in images (*e.g.* SIFT [142], SURF [143], or edge features [144]), and the geometric relationships between distinct features can then be used to compare images for registration purposes, or used to calculate pose directly. Pixel or feature-based registration may also be complemented with learning based tools such as Neural Networks (*e.g.* [145]). A detailed discussion of image matching and similarity metrics is given in Section 3.2.2.

Harris' template-matching method for articulation angle sensing [73] (Section 1.5) is an example of pixel-based matching for pose estimation, making use of cross-correlation.

Miao's dual camera ground-watching method for trailer off-tracking estimation [3] (1.6.1) is an example of feature-based matching for pose estimation, making use of SURF features.

1.8 Summary

- High Capacity Vehicles (HCVs) offer an attractive and feasible solution to reducing emissions and improving freight efficiency, and their uptake has grown following successful trials and implementations throughout the world.
- One of the barriers to increased uptake of HCVs, especially multiply-articulated HCVs, is their restricted low-speed manoeuvrability and potentially compromised stability at motorway speeds. A number of technologies such as path-following trailer steering, autonomous reversing, and jackknife control help to address these limitations.
- These technologies impose additional instrumentation and sensing requirements for trucks and trailers. In particular, the measurement of articulation angle and trailer off-tracking has been identified as critical to the functioning of these technologies.
- Existing articulation angle sensors are either trailer-based, require non-standard tractor-trailer communication links, require impractical trailer modifications or markers, or are not sufficiently accurate. Ideally the sensor should be purely tractor-based in order for it to be commercially feasible and should yield errors not exceeding 0.6°, with RMS errors below 0.4°.
- Existing methods of trailer off-tracking estimation are limited to high friction conditions and require tractor parameter information, both of which are not ideal. Miao's 'ground-watching' concept [3] addresses this friction limitation but presents additional practical limitations in off-highway conditions. An ideal trailer off-tracking sensor should yield errors not exceeding 0.15 m, with RMS errors below 0.10 m.
- Vision-based methods for pose and ego-motion estimation—including SLAM, visual odometry and template-matching—have proven effective and versatile in comparable sensing applications. Cameras are also a cost effective and commercially feasible technology compared to alternatives such as lidar and radar.
- In this work, novel vision-based concepts for articulation angle sensing and trailer off-tracking estimation are presented which address many of the limitations of the current state-of-the-art, thereby improving the commercial prospects of important HGV control systems. Furthermore, the concepts are validated in experiments on full-scale HGV combinations under various conditions, giving reliable evidence as to their effectiveness in real commercial applications.

1.9 Objectives

1. Develop a tractor-based articulation angle sensing concept which:
   (a) requires minimal information exchange with the trailer,
   (b) can work with any trailer or dolly, without significant modification,
   (c) is compatible with multiple truck and trailer combinations,
   (d) can provide angle measurements at 10 Hz with a maximum error of 0.6° and an RMS error of less than 0.4°.
2. Evaluate the performance of the articulation angle sensor in full-scale vehicle tests.
3. Develop a trailer-based trailer off-tracking estimation concept, which:
   (a) requires minimal information exchange with the truck/tractor,
   (b) makes no assumption about the state of wheel slip on any of the axles,
   (c) can provide off-tracking measurements at 10 Hz with a maximum error of 0.15 m and an RMS error of less than 0.10 m.
4. Evaluate the performance of the off-tracking sensor in full-scale vehicle tests.

1.10 Thesis structure

Chapter 2 provides a brief introduction to the topic of computer vision, to aid the reader as these concepts are used and referenced in later chapters. Thereafter, this thesis is primarily divided into two distinct parts: articulation angle sensing and trailer off-tracking estimation.

Chapter 3 covers the development of a vision-based articulation angle sensing concept, and investigates its feasibility and performance through simulations. Template-matching and PTAM-based algorithms are investigated. In Chapter 4, the concept is implemented on a full-scale tractor semi-trailer combination, and field testing results are presented and discussed. The concept is extended in Chapter 5 to accommodate a truck and full-trailer combination, where two articulation angles measurements are required, and field testing results on a full-scale vehicle combination are presented.

The investigation into trailer off-tracking estimation begins in Chapter 6. A vision-based concept using stereo visual odometry is introduced and evaluated in simulations. Chapter 7 presents the results of field tests on a tractor semi-trailer combination, wherein the performance of the concept is evaluated, and scope for future work identified.

Chapter 8 concludes the thesis with a summary of the main findings of each chapter, and a discussion of recommendations for future work.

1.11 Publications

The following peer-reviewed publications have resulted from this work [146, 147]:

1. C. de Saxe, D. Cebon, "A Visual Template-Matching Method for Articulation Angle Measurement," IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas, 2015, pp. 626-631, doi: 10.1109/ITSC.2015.108

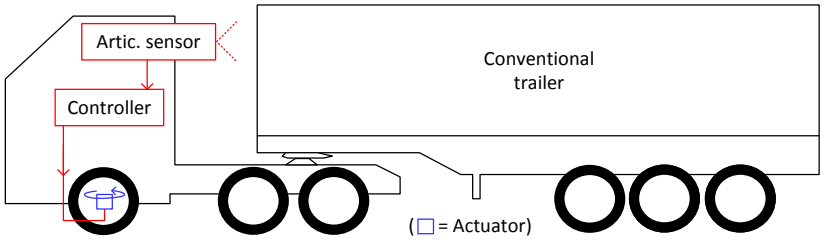
2. C. de Saxe, D. Cebon, "Visual odometry for trailer off-tracking estimation", 14th International Symposium on Heavy Vehicle Transport Technology, Rotorua, 2016, pp. 138-149, ISBN: 978-0-620-74483-6

In addition, three academic papers are in preparation to be submitted for publication in peer-reviewed journals. Broadly speaking, these have been categorised according to the themes of: articulation angle sensing using template-matching, articulation angle sensing using PTAM, and off-tracking sensing using visual odometry.

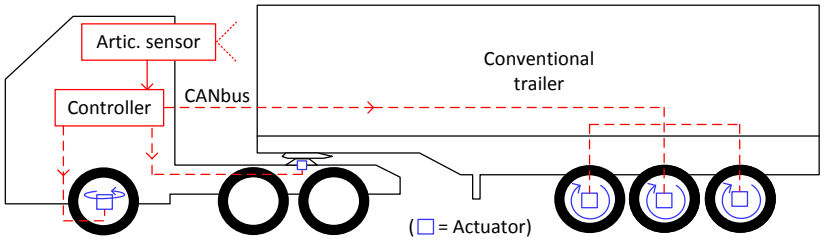
1.12 Tables and figures

Table 1.1 Summary of existing articulation angle estimation work

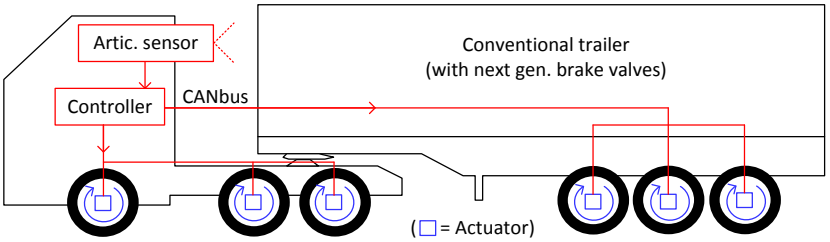
Source	Method	Validation	Γ_{max} (°)	ε_{max} (°)	Notes
Bouteldja [30]	State observer	Sim.	90	8	(Yaw angle)
Chu [52]	State observer	Sim.	3–20	0.3–10	
Ehlgen [70]	State observer	Veh.	48	5.4	
Schikora [71]	Infrared vision	Veh.	Unknown	Unknown	Markers
Schikora [71]	Encoder plate	Veh.	90	0.3	
Caup [72]	Vision (templates)	Veh.	20–55	5.5–7.6	
Harris [73]	Vision (hom. dec.)	Sim.	17–90	3.2–8.4	Planar trailer
Harris [73]	Vision (stereo)	Veh.	17-52	3.3–18	Planar trailer
Fuchs [74]	Vision	Sim.	50	0.5	Markers
Fuchs [75]	Vision+KF	Sim.	15–30	0.6	Markers



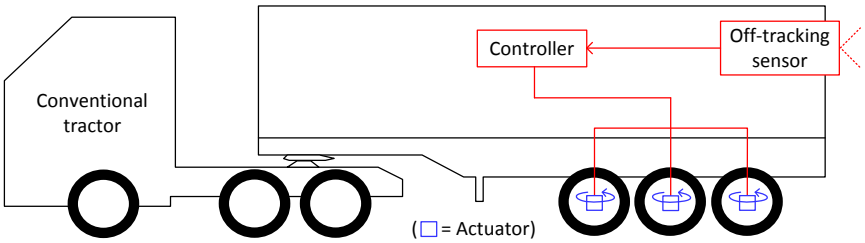
(a) Tractor-based control for autonomous reversing



(b) Tractor-based control for jackknife prevention

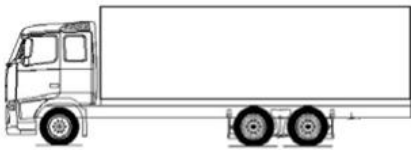


(c) Tractor-based control for combined braking and steering

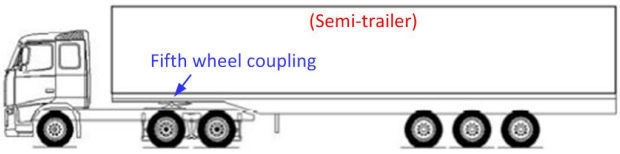


(d) Trailer-based control for trailer steering

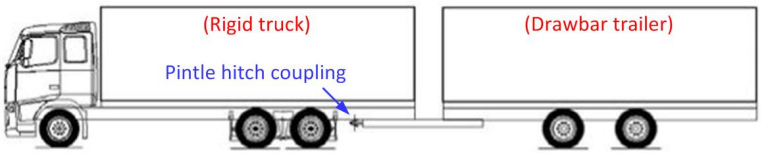
Fig. 1.1 Tractor- and trailer-based control ideologies



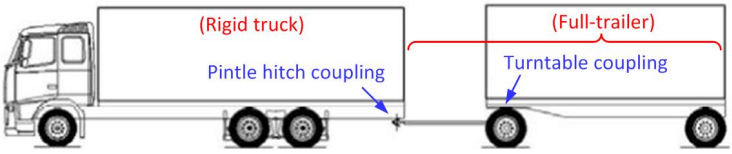
(a) Rigid truck



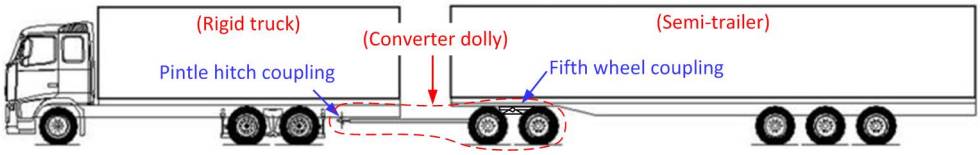
(b) Tractor semi-trailer



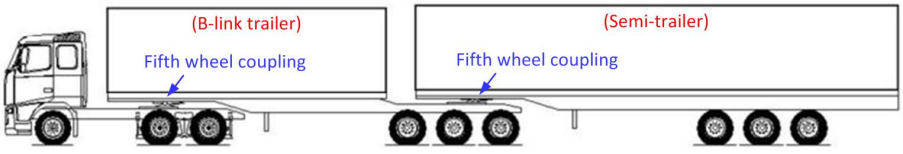
(c) Truck and drawbar trailer (also 'centre-axle trailer'/'dog trailer')



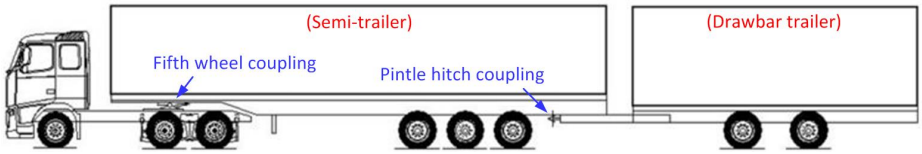
(d) Truck and full-trailer (also 'turntable drawbar trailer')



(e) Truck, converter dolly and semi-trailer ('Nordic combination')

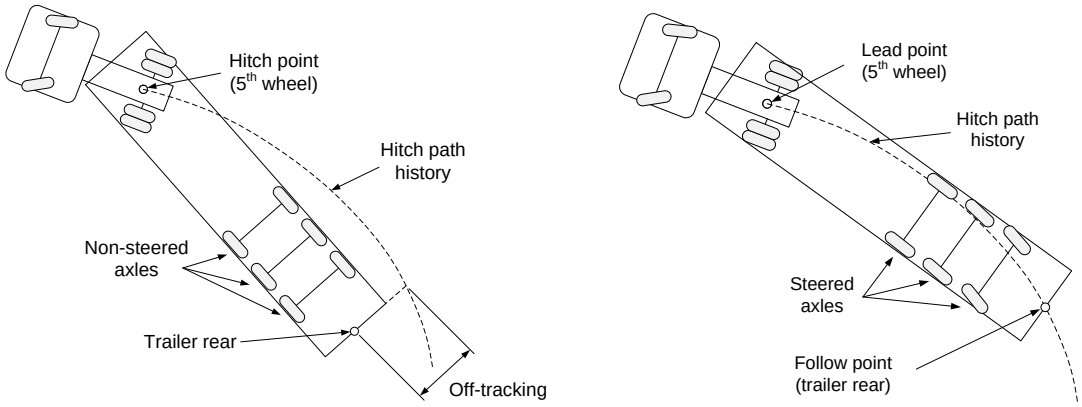


(f) B-double



(g) Tractor, semi-trailer and drawbar trailer

Fig. 1.2 Common HGV and HCV combinations (illustrations from [1])



(a) (b)

Fig. 1.3 Trailer off-tracking: (a) unsteered trailer, (b) steered trailer with path-following control

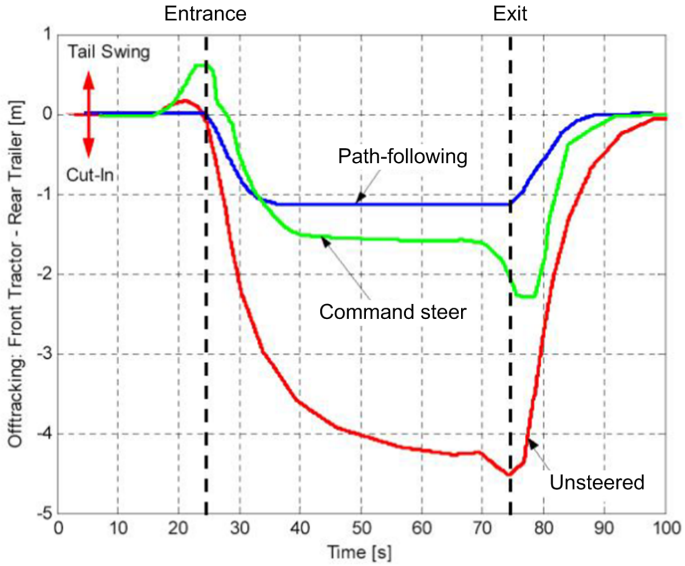


Fig. 1.4 Trailer steering performance in a roundabout manoeuvre (modified from [2]).

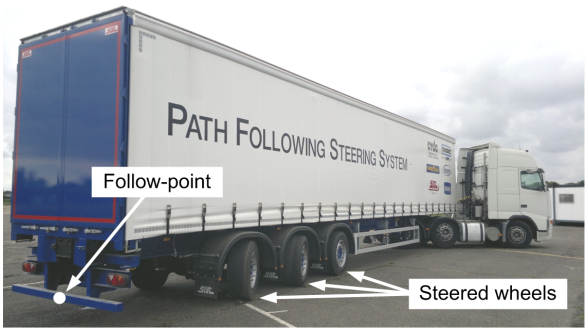
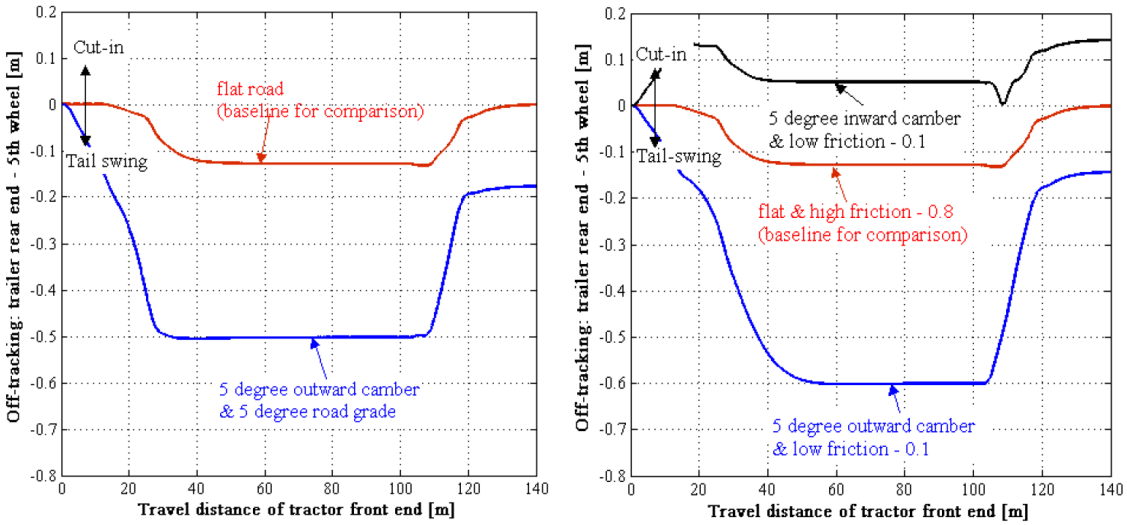


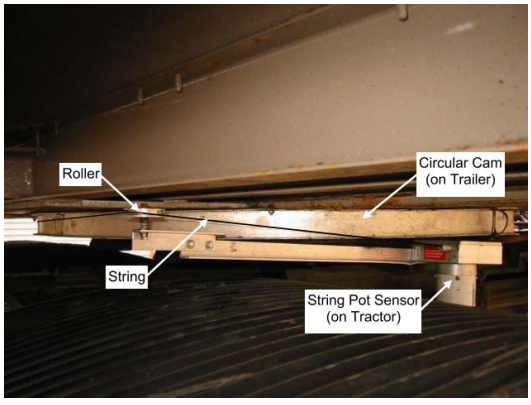
Fig. 1.5 CVDC's prototype path-following semi-trailer



(a) Road grade and camber (b) Road camber and reduced friction

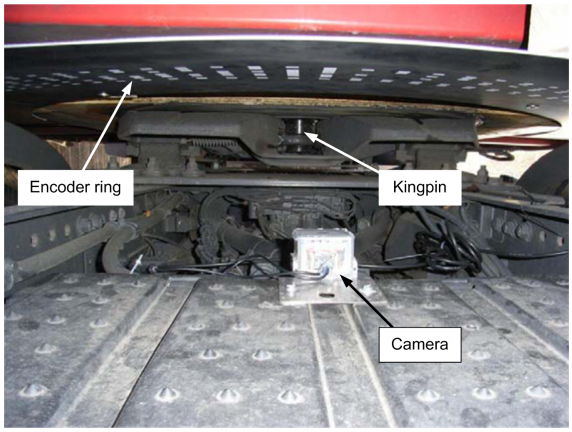
Fig. 1.6 CVDC path-following performance under adverse road conditions [3]

(a) VSE kingpin sensor [66] (b) Orbisense kingpin sensor [28]

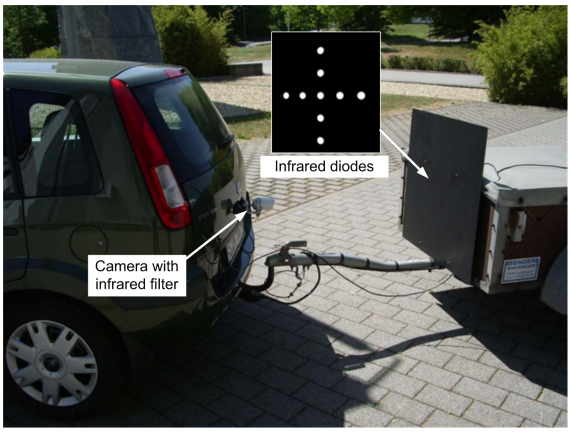


(c) String potentiometer sensor [28]

Fig. 1.7 Commercial and research articulation angle sensors.

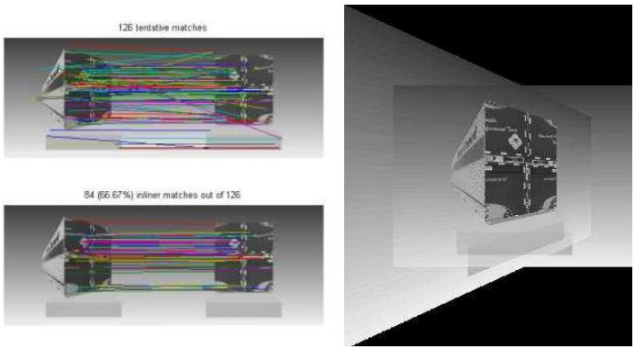


(a) Schikora, semi-trailer [71]

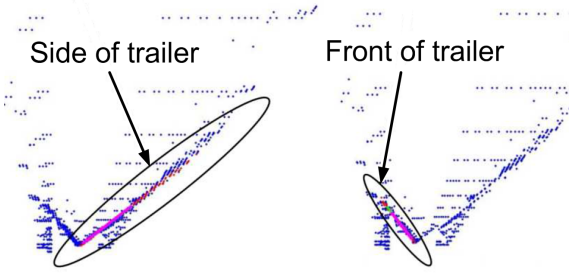


(b) Schikora, drawbar trailer [71]

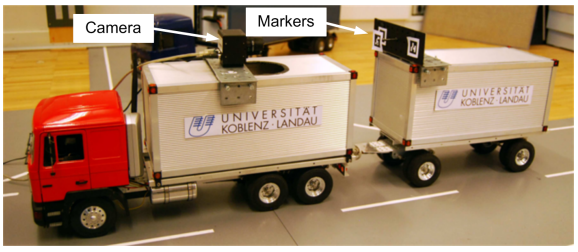
(c) Caup [72]



(d) Harris, homography decomposition [73]

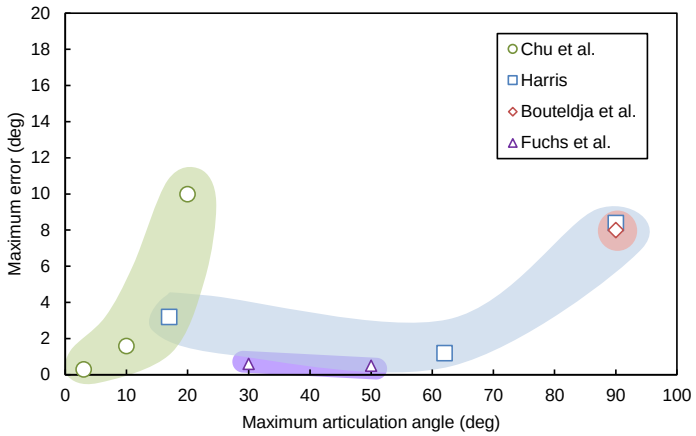


(e) Harris, stereo [73]

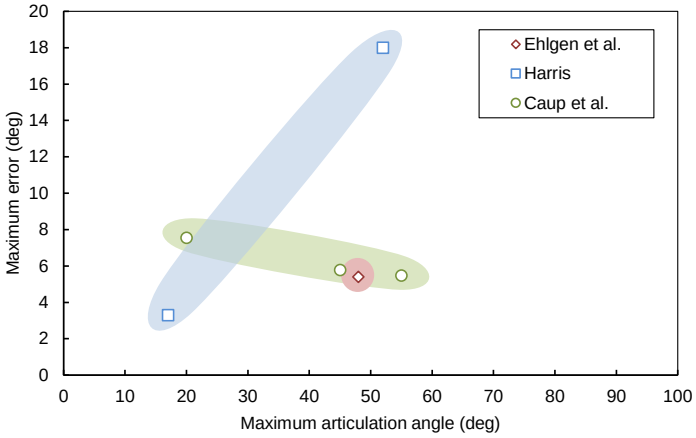


(f) Fuchs [74]

Fig. 1.8 Non-contact articulation angle sensors



(a) Results based on simulations



(b) Results based on vehicle tests

Fig. 1.9 Performance summary of published articulation angle sensing solutions, showing maximum errors obtained as a function of maximum articulation angles assessed

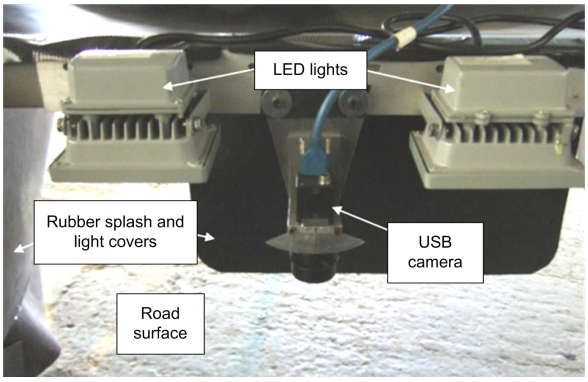
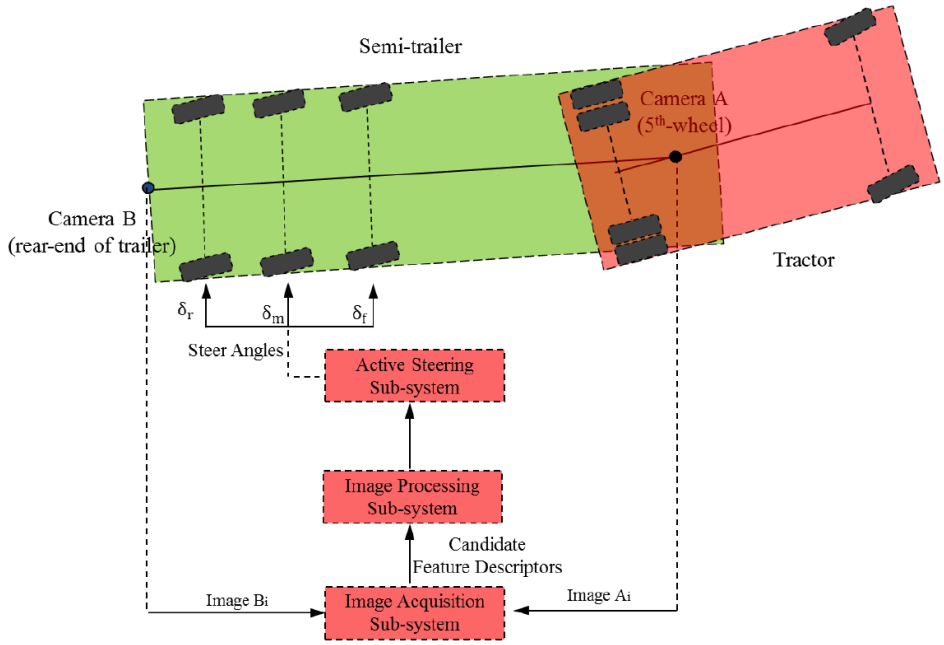
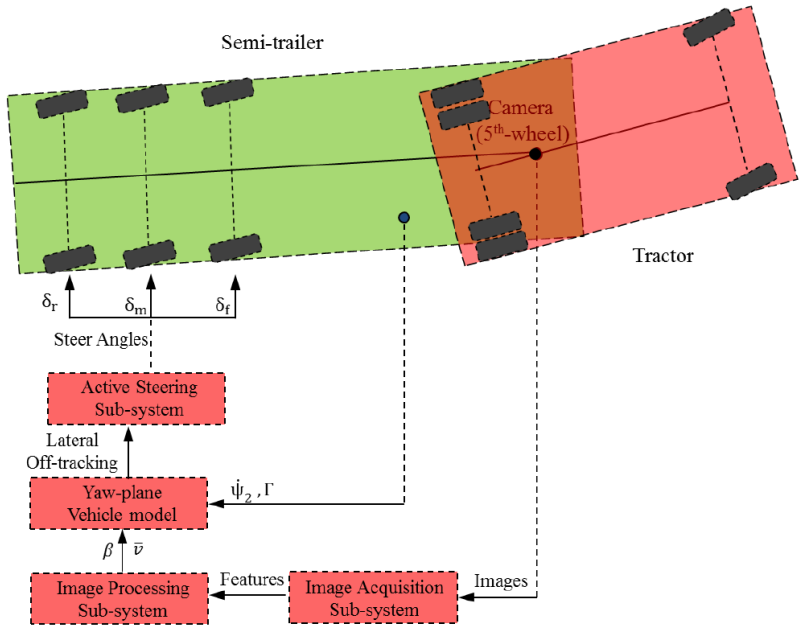


Fig. 1.10 Miao's 'ground-watching' camera (adapted from [3])



(a) DC-GWNS



(b) SC-GWNS

Fig. 1.11 Miao's dual (a) and single camera (b) ground-watching navigation systems [3]

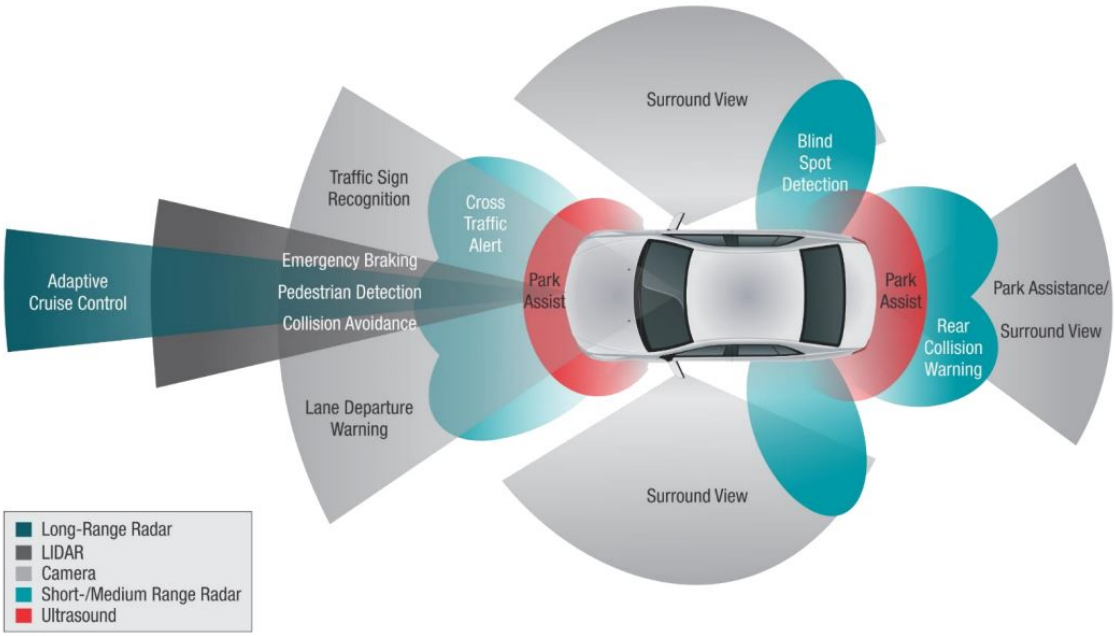


Fig. 1.12 Typical arrangement sensors on a modern passenger vehicle [4]

(a) (b)

Fig. 1.13 Google's self-driving car (a), with Velodyne HDL-64E S2 LiDAR system (b) [5]

Chapter 2

Computer vision fundamentals

This brief chapter has been included to provide an overview of relevant vision-related theory and terminology, for readers without a background in computer vision and also to lay the foundation for concepts and theory used in subsequent chapters. 'Computer vision' refers to a collection of techniques and algorithms used for the efficient and useful extraction of information from image data. Such data may be obtained from a single image, from simultaneous images taken by two or more cameras (*e.g.* stereo vision), or from an image sequence. The use of computer vision for pose and motion estimation tasks usually requires two fundamental steps: extracting points of interest from an image or images ('feature detection and description'), and transforming the locations of these interest points from image space into real world coordinates ('triangulation'). Conversely, the mapping of real world points into image space is known as 'perspective projection'.

2.1 Images, features and descriptors

An image consists of a 2-D array of pixels, each with an associated greyscale intensity or 3-value RGB colour value (Red Green Blue). In this work only greyscale images are considered, which commonly utilise an 8-bit representation with pixel values between 0 (black) and 255 (white). The locations and intensities of pixels or groups of pixels are used by computer vision algorithms to perform detection or motion estimation tasks.

Computer vision techniques are usually classified as either 'dense' or 'sparse'. Dense techniques utilise raw pixel intensities over the full image (or smaller 'patches' extracted from the image) to detect objects, motion, or structure. Sparse techniques add an intermediate step in which distinct points of interest are first detected in the image. These are referred to as 'features' and often have an associated 'descriptor' to aid in feature tracking and feature matching between images.

Features are usually detected based on common groupings of pixel intensities, or particular discontinuities in pixel intensity gradients. Common image features include edges [144, 148], corners [149] and 'blobs' [148]. Edges are not well-suited to motion detection, but can be used to identify shapes (using Hough transforms for example [150]). Corners and blobs are more useful for motion detection.

The purpose of a 'feature detector' is to locate features in an image. Convolving an image with a Gaussian kernel, Laplacian of Gaussian (LoG) kernel, or a Difference of Gaussian (DoG) kernel is often used to detect peaks, troughs or zero-crossings of pixel intensity gradients, from which various features are defined. Traditional feature detectors include the Canny edge detector [144], the Harris corner detector [149], and the DoG blob detector [151]. A popular and efficient corner detector is the Features from Accelerated Segment Test (FAST) [152] in which the variation in pixel intensities in a circle around a point are used to determine whether or not the point is a corner.

A 'feature descriptor' is a type of 'fingerprint' associated with a feature which enables features to be tracked and matched between images. The simplest descriptor is a patch of pixels around the feature, which can be tracked or matched through cross correlation techniques. More advanced descriptor types include SIFT (Scale-Invariant Feature Transform) [142], SURF (Speeded Up Robust Features) [143], HOG (Histogram of Oriented Gradients) [153], GLOH (Gradient Location and Orientation Histogram) [154] and wavelets [155].

The SIFT feature descriptor has been used extensively in the literature. It uses a 4×4 grid of cells around a keypoint, each consisting of 4×4 pixels. At each pixel the intensity gradient magnitude and direction is found and grouped into one of 8 directions, and these are grouped into one histogram of gradients per cell. The result is a $4 \times 4 \times 8$ dimension descriptor vector per feature. Intensity gradients are found with the DoG operator. The SURF feature descriptor is a more efficient development of the SIFT feature through the use of 'box filters', and has been adopted in place of SIFT descriptors in many applications that require computational efficiency.

Matching features between two or more images is a useful way to determine structure or motion. If the images are captured by two cameras simultaneously (stereo vision), metric structure can be extracted from a scene using triangulation. If the images are taken sequentially with the same camera, the motion of the camera and/or the structure of the scene can be calculated, but usually only up to an arbitrary scale (*i.e.* depth calculation is not possible). Computer vision techniques utilising these concepts to infer structure or motion include 'visual odometry' (see [82]), 'optical flow' (*e.g.* [156]) and 'structure from motion' (*e.g.* [133]).

2.2 Perspective projection

Perspective projection is a model of the imaging process in which points in 3-D space are transformed into coordinates in an image. The theory of this imaging process can be described using a simplified 'pin-hole' camera model. In the pin-hole model (Figure 2.1), light rays from 3-D world points coincide at the camera's focal point (taken to be the origin of the camera co-ordinate system), passing through the virtual imaging plane located at a distance f (the focal length) ahead of the camera. Co-ordinates of points on the imaging plane are then related to pixel co-ordinates through a simulated CCD (Charge-Coupled Device) imaging process.

The pin-hole imaging process comprises a sequence of transformations from world coordinates $\mathbf{X} = (X,Y,Z)$, to camera-centred coordinates $\mathbf{X}_c = (X_c,Y_c,Z_c)$, to image plane coordinates $\mathbf{X}_I = (X_I,Y_I)$, and finally to pixel coordinates in the image plane $\mathbf{w} = (u,v)$. These transformations will first be outlined individually, and then brought together in the full pin-hole camera model. The perspective projection theory presented here is adapted from [157] and [158].

The rigid body transformation from world coordinates $\mathbf{X}$ to camera-centred coordinates $\mathbf{X}_c$ is governed by a rotation R and translation $\mathbf{T}$ as follows:

$$\begin{aligned} \mathbf{X}_c &= \mathrm{R}\mathbf{X} + \mathbf{T} \\ \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} &= \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \end{aligned} \tag{2.1}$$

The 'perspective projection transformation' from camera-centred coordinates $\mathbf{X}_c$ to coordinates in the image plane $\mathbf{X}_I$ is a function of the camera focal length f. The transformation is as follows:

$$X_I = f_x \frac{X_c}{Z_c} \qquad Y_I = f_y \frac{Y_c}{Z_c} \tag{2.2}$$

where f_x and f_y are the focal lengths in the horizontal and vertical directions.

Finally, the 'CCD imaging transformation' from image plane coordinates $\mathbf{X}_I$ to pixel coordinates $\mathbf{w}$ is given by:

$$u = u_0 + k_u X_I \qquad v = v_0 + k_v Y_I \tag{2.3}$$

where (u_0,v_0) is the principal point (the location at which the optical axis passes through the image plane, usually near the geometric centre of the image), and k_u and k_v are scaling parameters which account for rectangular pixels.

It is common to describe world, camera, image plane and pixel co-ordinates in homogeneous form to linearise the imaging process and simplify subsequent operations. The coordinates in homogeneous form are:

$$\begin{aligned} \tilde{\mathbf{X}} &= \begin{bmatrix} \xi X & \xi Y & \xi Z & \xi \end{bmatrix}^\mathsf{T} & \tilde{\mathbf{X}}_c &= \begin{bmatrix} \xi X_c & \xi Y_c & \xi Z_c & \xi \end{bmatrix}^\mathsf{T} \\ \tilde{\mathbf{X}}_I &= \begin{bmatrix} \xi X_I & \xi Y_I & \xi \end{bmatrix}^\mathsf{T} & \tilde{\mathbf{w}} &= \begin{bmatrix} \xi u & \xi v & \xi \end{bmatrix}^\mathsf{T} \end{aligned} \tag{2.4}$$

where the tilde symbol denotes homogeneous co-ordinates and ξ is an arbitrary scaling factor. Homogeneous coordinates are useful in computer vision in that they enable points at infinity to be represented by finite coordinates, and they enable simple matrix representation of important image transformation operations. Wherever possible, ξ is set to 1, but can otherwise take on any value.[1]

In homogeneous form, Equations 2.1, 2.2 and 2.3 become:

$$\tilde{\mathbf{X}}_c = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tilde{\mathbf{X}} \tag{2.5}$$

$$\tilde{\mathbf{X}}_I = \begin{bmatrix} f_x & 0 & 0 & 0 \\ 0 & f_y & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tilde{\mathbf{X}}_c \tag{2.6}$$

$$\tilde{\mathbf{w}} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tilde{\mathbf{X}}_I \tag{2.7}$$

[1] In converting Cartesian to homogeneous co-ordinates, the choice of ξ is arbitrary as this is a one-to-many mapping (homogeneous co-ordinates are scale-invariant). It is therefore simplest to set $\xi = 1$. From homogeneous to Cartesian co-ordinates, ξ will not necessarily be equal to 1, and homogeneous co-ordinates must be divided through by ξ to yield the correct Cartesian co-ordinates.

The homogeneous form of these equations enables the overall imaging process to be simplified to the following form:

$$\tilde{\mathbf{w}} = \mathrm{K}\left[\begin{array}{c|c} \mathrm{R} & \mathbf{T} \end{array}\right]\tilde{\mathbf{X}}$$

$$\begin{bmatrix} \xi u \\ \xi v \\ \xi \end{bmatrix} = \begin{bmatrix} k_u f_x & 0 & u_0 \\ 0 & k_v f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2.8}$$

where K is the camera calibration matrix containing the 'intrinsic' camera parameters. ξ is arbitrarily set to 1 in the world point for simplicity, but will have a non-unit value in $\tilde{\mathbf{w}}$.

Given the pixel coordinates of an image feature and the intrinsic camera parameters, Equation 2.8 may be used to determine the coordinates of that point in 3-D space, *but only up to an arbitrary scale factor due to* ξ. In other words, using a single camera one can determine the ray along which the point lies in space, but not the point's distance from the camera. This is sometimes referred to as 'scale ambiguity'.

To remove the scale ambiguity in the imaging process, triangulation is required using a second image of the point from an alternative viewpoint. The rotation and translation between the viewpoints must be known. This is the basis of stereo vision.

2.2.1 Distortion correction

The pinhole camera model described by Equation 2.8 assumes that straight lines in world coordinates remain straight in the image plane. In reality, camera lenses introduce non-linear distortion to the imaging process, and the effects of this distortion must be removed before the pinhole camera model is applicable.

The distortion model described by Heikkilä [159] incorporates both radial and tangential distortion. In Heikkilä's model, the effects of radial distortion are modelled by:

$$\begin{aligned} X_{I,\text{distorted}} &= X_I(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \ldots) \\ Y_{I,\text{distorted}} &= Y_I(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \ldots) \end{aligned} \tag{2.9}$$

where $r = \sqrt{u^2 + v^2}$ and k_1, k_2, ... are the radial distortion coefficients. It is common for two or three radial distortion coefficients to be adopted, but additional coefficients can be

used for more severe distortions. Tangential distortion is modelled as follows:

$$\begin{aligned} X_{I,\text{distorted}} &= X_I + 2p_1 X_I Y_I + p_2(r^2 + 2X_I^2) \\ Y_{I,\text{distorted}} &= Y_I + 2p_2 X_I Y_I + p_1(r^2 + 2Y_I^2) \end{aligned} \tag{2.10}$$

where p_1 and p_2 are the tangential distortion coefficients.

Often a simplified distortion model is adopted, such as the 'FOV-model' of Devernay and Faugeras [160] which utilises a single radial distortion coefficient. This simplified model is used in Klein and Murray's PTAM [123]. In this case the transformations from camera-centred coordinates to pixels (Equations 2.6 and 2.7) are modified as follows:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \frac{r'_{\text{PTAM}}}{r_{\text{PTAM}}} \begin{bmatrix} X_c/Z_c \\ Y_c/Z_c \end{bmatrix} \tag{2.11}$$

where

$$r_{\text{PTAM}} = \sqrt{\frac{X_c^2 + Y_c^2}{Z_c^2}} \qquad r'_{\text{PTAM}} = \frac{1}{s}\arctan\left(2r_{PTAM}\tan\frac{s}{2}\right) \tag{2.12}$$

and s is the distortion coefficient.

Equations 2.9 and 2.10 (or Equations 2.11 and 2.12) can be incorporated into the pinhole camera model (Equation 2.8) to model the full distorted imaging process. In practice, distortion is usually corrected first, and then the pinhole camera model is assumed valid in subsequent operations for simplicity.

2.3 Camera calibration

The intrinsic camera parameters, f_x, f_y, k_u, k_v, u_0 and v_0 can be obtained through a 'camera calibration' process such as described by Zhang [161]. If Equation 2.8 is updated to include distortion, the distortion parameters k_1, k_2, ... and p_1, p_2 can be determined at the same time.

Given a number of known solutions for $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{X}}$, optimization techniques may be used to solve for K, R, $\mathbf{T}$ in Equation 2.8. In practice, this typically involves capturing images of a known checkerboard pattern at multiple orientations and locations relative to the camera. For each image, corner detection methods can be used to detect the checkerboard corners and hence determine $\tilde{\mathbf{w}}$. World coordinates of the pattern corners $\tilde{\mathbf{X}}$ are known given the size and number of squares in the checkerboard grid, and by defining the origin of the world coordinates to be coincident with one of the corners of the checkerboard.

Popular implementations of the camera calibration process include the MATLAB 'Computer Vision System Toolbox' [162], Bouguet's 'Camera calibration toolbox for MATLAB' [163] and the OpenCV camera calibration functions [164]. All these implementations utilise Zhang's calibration process [161] with Heikkilä's distortion model [159]. The MATLAB implementations utilise three radial distortion coefficients (k_1–k_3) while the OpenCV implementation utilises up to six (k_1–k_6).

The 'reprojection error' is defined as the error between the observed pixel location of the point $\tilde{\mathbf{w}}$ and the predicted pixel location projected onto the image plane from the world coordinate $\tilde{\mathbf{X}}$ using Equation 2.8. The goal of the optimization algorithm is to minimise the squared sum of reprojection errors in each image. Usually between 20 and 30 images are required for the optimization to converge to acceptable accuracy (≈ 0.5 pixels).

2.4 Stereo camera calibration

A stereo camera pair consists of two cameras (left and right) separated by a fixed distance (the 'stereo baseline'), capturing simultaneous images of a scene from different viewpoints. Using feature-matching and triangulation, stereo cameras are able to determine depth information from a scene, thereby removing the scale ambiguity inherent in mono camera systems.

Traditional stereo vision algorithms (see [157]) require that the relative rotation and translation between the left and right cameras must be known, in order to perform triangulation. It is sometimes useful, but not necessary, to also ensure that the optical axes are parallel and that the left and right image planes are coplanar. This yields horizontal epipolar lines, meaning that left-right feature matching reduces to a simple horizontal scanning process [157]. This is approximated by mounting the cameras laterally aligned and in parallel (parallel stereo), but small misalignments must be compensated for. A 'stereo rectification' process is used to apply small corrections to left and right images to compensate for these misalignments.

In general, the stereo baseline and relative orientation between the left and right cameras (the 'extrinsic' camera parameters) will not be known precisely and must be determined through a 'stereo calibration' process. This is in addition to the single camera calibration process required to determine the 'intrinsic' camera parameters. Once the relative orientation of the cameras is known, 'stereo rectification' may be performed. Distortion correction is usually applied to the images first, followed by stereo rectification. In this work, parallel stereo is assumed, and so stereo rectification is required.

The goal of the stereo calibration process is to estimate the rotation and translation from right to left camera origins, denoted R_{rl} and $\mathbf{T}_{rl}$ (where l and r refer to the left and right cameras). With reference to Figure 2.2, for a common point in 3-D space observed in both

left and right images, the camera-centred coordinates of that point are related as follows:

$$\mathbf{X}_{c,r} = \mathrm{R}_{rl}\mathbf{X}_{c,l} + \mathbf{T}_{rl} \tag{2.13}$$

From this, using Equation 2.1, it can be shown that:

$$\mathrm{R}_{rl} = \mathrm{R}_r\mathrm{R}_l^\mathsf{T} \tag{2.14}$$

$$\mathbf{T}_{rl} = \mathbf{T}_l - \mathrm{R}_{rl}^\mathsf{T}\mathbf{T}_r \tag{2.15}$$

In a similar manner to the single camera calibration process, the values for R_{rl} and $\mathbf{T}_{rl}$ can be estimated by observing a known checkerboard pattern ($\tilde{\mathbf{X}}$), and matching common feature points between left and right images ($\tilde{\mathbf{w}}_l$, $\tilde{\mathbf{w}}_r$). By observing the checkerboard from multiple viewpoints, Equations 2.8, 2.14 and 2.15 can be solved for R_{rl} and $\mathbf{T}_{rl}$ using a least squares optimization scheme, minimising the reprojection errors.

Once a stereo camera rig has been calibrated, the rectification transformations must be computed so that images can be rectified in real-time. The following is a simplified overview of the rectification process described by Fusiello *et al.* [165].

Rectification involves applying rotation transformations to left and right images to align the optical axes, followed by a scale adjustment of one of the images to align the image planes. The rectification rotation matrix is calculated as:

$$\mathrm{R}_{rect} = \begin{bmatrix}\mathbf{e}_1^\mathsf{T} & \mathbf{e}_2^\mathsf{T} & \mathbf{e}_3^\mathsf{T}\end{bmatrix}^\mathsf{T} \tag{2.16}$$

where:

$$\mathbf{e}_1 = \frac{\mathbf{T}_{rl}}{\|\mathbf{T}_{rl}\|} \qquad \mathbf{e}_2 = \frac{1}{\sqrt{T_{rl,x}^2 + T_{rl,y}^2}}\begin{bmatrix}-T_{rl,y} & T_{rl,x} & 0\end{bmatrix}^\mathsf{T} \qquad \mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2 \tag{2.17}$$

Left and right camera coordinates can then be rectified as follows:

$$\mathbf{X}_{c,l,rect} = \mathrm{R}_{rect}\mathbf{X}_{c,l} \qquad \mathbf{X}_{c,r,rect} = \mathrm{R}_{rl}\mathrm{R}_{rect}\mathbf{X}_{c,r} \tag{2.18}$$

followed by a simple scale correction.

Implementations of the stereo calibration and rectification algorithms based on the above theory are available in MATLAB's 'Computer Vision System Toolbox' [162] and in the OpenCV libraries for C++ and Python [164].

2.5 Figures

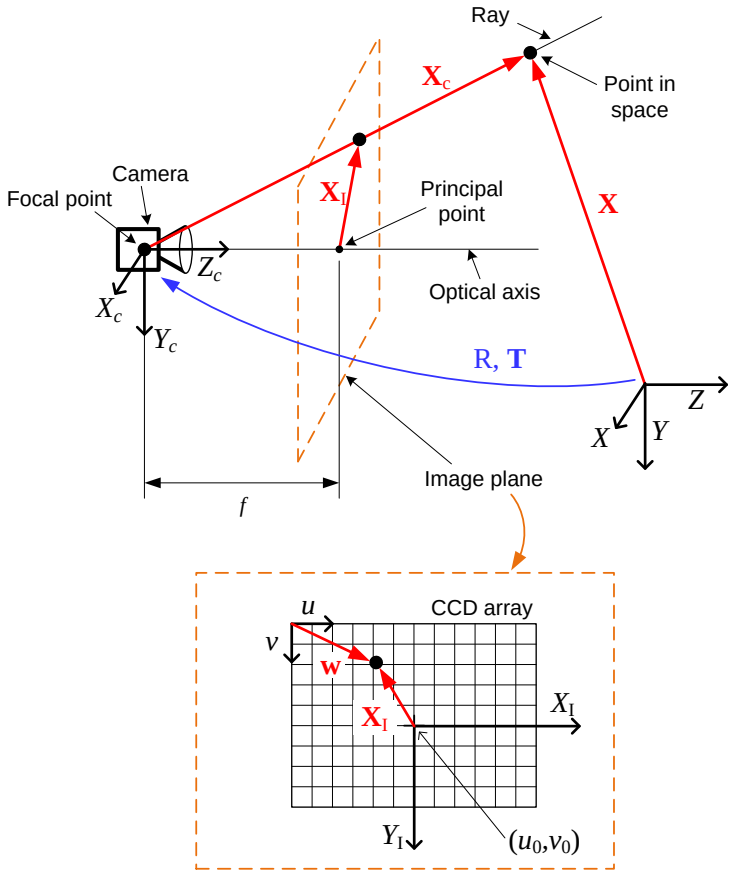


Fig. 2.1 Perspective projection for a pin-hole camera model

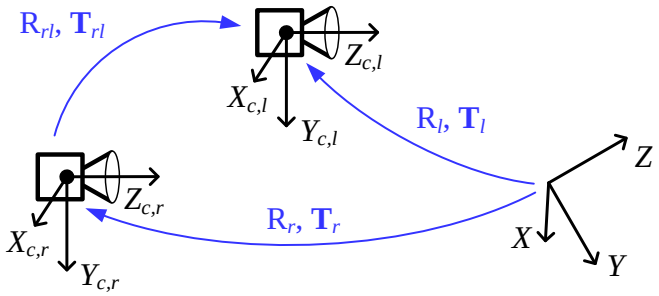


Fig. 2.2 Rotation and translation transformations in stereo vision

Chapter 3

Development of an articulation angle sensor

In this chapter, two vision-based methods for measuring the articulation angle of articulated HGVs are described: template-matching and Parallel Tracking and Mapping (PTAM). A single rear-facing camera centrally mounted behind the tractor cab was adopted in a similar manner to previous work on the topic by Harris [73]. Simulations of the proposed methods were conducted using a model of a tractor semi-trailer, comprising a dynamic 'bicycle model' for vehicle dynamics and a CAD-based visual model to generate image data.

The chapter begins with description of the vehicle simulation model used. Next, the theory and implementation of the template-matching and PTAM methods are described, followed by a simulation-based investigation into their performance and behaviour.

3.1 Simulation model

The simulation environment consisted of a tractor semi-trailer vehicle model and a CAD-based visual model. The vehicle model was used to generate appropriate articulation angle signals for a given steer input and vehicle speed. The visual model was used to generate representative image data to match these articulation angle signals.

3.1.1 Dynamic vehicle model

A dynamic 'bicycle model' of a tractor semi-trailer was used to simulate vehicle motion. The model had three degrees of freedom: lateral velocity, yaw and articulation. Longitudinal speed was assumed to be constant. A free-body diagram of the model is shown in Figure 3.1.

Variables m and J are vehicle mass and moment of inertia about the centre of gravity (CoG); U and V are longitudinal and lateral velocities; Ω is yaw angular velocity; $F_{h,x}$ and $F_{h,y}$ are longitudinal and lateral hitch forces (measured in the tractor unit's coordinate system); F_f, F_r and F_t are tractor front, tractor rear and trailer lateral tyre forces; and δ is the steer angle at the front axle. Γ is the articulation angle. Subscripts 1 and 2 denote the tractor and trailer units respectively.

The following assumptions were made:

1. Left and right tyres may be reduced to a single tyre with common steer angle at the axle centre.
2. Pitch and roll motion is small and can be neglected.
3. Lateral tyre forces are linearly related to tyre sideslip angles by a constant coefficient, C, the cornering stiffness.
4. The forward speed of the tractor, U_1, is constant.

From Figure 3.1, balancing lateral forces acting on the tractor and trailer units yields:

$$m_1\left(\dot{V}_1 + U_1\Omega_1\right) + F_f\cos\delta + F_r - F_{h,y} = 0 \tag{3.1}$$

$$m_2\left(\dot{V}_2 + U_2\Omega_2\right) + F_t + F_{h,y}\cos\Gamma + F_{h,x}\sin\Gamma = 0 \tag{3.2}$$

Balancing yaw moments about the hitch point for the tractor and trailer units yields a further two equations of motion:

$$(a_1 + c_1)F_f\cos\delta + c_1 m_1(\dot{V}_1 + U_1\Omega_1) + J_1\dot{\Omega}_1 - (b_1 - c_1)F_r = 0 \tag{3.3}$$

$$(a_2 + b_2)F_t + a_2 m_2\left(\dot{V}_2 + U_2\Omega_2\right) - J_2\dot{\Omega}_2 = 0 \tag{3.4}$$

Tyre forces are related to slip angles by a linear cornering stiffness, C, as follows:

$$F_f = C_f\alpha_f \qquad F_r = C_r\alpha_r \qquad F_t = C_t\alpha_t \tag{3.5}$$

where C_f, C_r, and C_t are the sums of front, rear and trailer cornering stiffnesses respectively. The respective tyre side-slip angles are

$$\begin{aligned}
\alpha_f &= \tan^{-1}\left(\frac{V_1 + a_1\Omega_1}{U_1}\right) - \delta \\
\alpha_r &= \tan^{-1}\left(\frac{V_1 - b_1\Omega_1}{U_1}\right) \\
\alpha_t &= \tan^{-1}\left(\frac{V_2 - b_2\Omega_2}{U_2}\right)
\end{aligned} \tag{3.6}$$

Trailer kinematics may be related to tractor kinematics using the constraint at the hitch point. Yaw rates Ω, lateral velocities V, and longitudinal velocities U, along with their first derivatives, are related as follows:

$$\Omega_2 = \Omega_1 + \dot{\Gamma} \tag{3.7}$$

$$\dot{\Omega}_2 = \dot{\Omega}_1 + \ddot{\Gamma} \tag{3.8}$$

$$V_2 = (V_1 - (b_1 - c_1)\Omega_1)\cos\Gamma - U_1 \sin\Gamma - a_2\Omega_2 \tag{3.9}$$

$$\dot{V}_2 = (\dot{V}_1 - (b_1 - c_1)\dot{\Omega}_1)\cos\Gamma - \dot{U}_1 \sin\Gamma - a_2\dot{\Omega}_2 - \dot{\Gamma}\left[(V_1 - (b_1 - c_1)\Omega_1)\sin\Gamma + U_1\cos\Gamma\right] \tag{3.10}$$

$$U_2 = U_1\cos\Gamma + (V_1 - (b_1 - c_1)\Omega_1)\sin\Gamma \tag{3.11}$$

$$\dot{U}_2 = \dot{U}_1\cos\Gamma + (\dot{V}_1 - (b_1 - c_1)\dot{\Omega}_1)\sin\Gamma + \dot{\Gamma}\left[(V_1 - (b_1 - c_1)\Omega_1)\cos\Gamma - U_1\sin\Gamma\right] \tag{3.12}$$

Equations 3.1 to 3.12 were implemented in MATLAB [166] and solved using a standard ordinary differential equation (ODE) solver. Inputs were tractor steer angle and longitudinal speed, and the output was articulation angle, Γ. Vehicle parameters were taken from [73] for a Volvo tractor unit and CVDC box-type semi-trailer. Details of the simulation parameters are given in Table 3.1.

3.1.2 CAD visual model

A simplified CAD model of a tractor and box-type semi-trailer combination was created in Autodesk Inventor [167]. The model is shown in Figure 3.2. The front and sides of the semi-trailer were assumed to be planar. The tractor unit was assumed fixed and the semi-trailer unit was free to rotate in the yaw-plane about the hitch point.

Representative visual texture was added to the trailer in order to provide the vision algorithms with a sufficient degree of visual detail. A lack of any visual detail would make any vision-based method ineffective. It is reasonable to assume some level of visual detail will be available on trailers, either through mechanical features such as bolts or joins, textural variations due to surface finish or wear, or company branding in the form of decals or paint detailing.

A virtual camera was located at the rear of the tractor cab[1] viewing the front of the semi-trailer in a similar configuration to that used by Harris [73]. The simulated camera in Autodesk Inventor functions as a simple 'pin-hole' camera which behaves according to basic

[1] In reality this means that the camera would be subjected to the motion of the cab, which is connected to the tractor chassis through a number of compliant suspension elements. The effect of cab motion relative to the chassis on articulation angle measurement is assumed to be small. In the CAD model, tractor cab and chassis are modelled as a single rigid body.

perspective projection theory (Equation 2.8) with no distortion. Ambient lighting was set to give a reasonable variation in lighting with articulation angle, but with no harsh shadows. The camera was set to capture greyscale images of VGA resolution (640×480 pixels). The camera was aligned so as to have its optical axis parallel with the longitudinal axis of the tractor (and hence the longitudinal axis of the trailer at $\Gamma = 0°$). However, it should be noted that the CAD software lacked the ability to perfect align the camera with a specified axis, and so the achieved alignment is only approximate.

Tractor wheelbase, hitch offset and trailer front overhang (h) were chosen to be representative of a typical tractor unit at 3.7 m, 0.775 m, and 1.2 m respectively (see Figure 3.2). The distance from the camera's optical centre to the front of the trailer (d) was chosen to be 2.3 m. This is larger than what would be expected given the other parameters (it suggests the rear of the tractor cab is ahead of the steer axle), but was chosen so as to provide a complete view of the trailer front throughout the full articulation angle range for the pinhole camera. The complete view of the trailer enabled the possibility of various other algorithms to be investigated, which could rely on extracting the trailer outline for example. Ultimately, such methods were deemed unfavourable, but the geometry was retained for the simulation studies. No significant effects were anticipated as a result of this, other than the trailer occupying a slightly smaller pixel space than in reality for a typical tractor semi-trailer combination.

The intrinsic parameters of the virtual camera in the CAD environment were determined without the need for a camera calibration process, given that: (i) there was no distortion, (ii) pixels were square, (iii) the optical axis was perfectly centred in each image, and (iv) the focal length was a direct function of the selected viewing angle (which can be set within the environment). Instead, these were calculated as follows:

$$k_u = k_v = 1 \qquad (u_0, v_0) = \left(\frac{n_u - 1}{2}, \frac{n_v - 1}{2} \right) \qquad f_x = f_y = f = \frac{n_u}{2 \tan\left(\frac{\text{FOV}}{2}\right)} \tag{3.13}$$

where n_u and n_v are the image width and height, and FOV is the field of view (adjustable within Autodesk Inventor). In this case: $f = 243$ pixels, $(u_0, v_0) = (320, 240)$ pixels.

3.2 Template-matching method

The underlying concept of the template-matching (TM) method is the ability to generate a predicted image of the front of the trailer at any articulation angle, using a reference view of the trailer (*i.e.* at $\Gamma = 0°$), knowledge of the parameters d and h, and assuming a planar trailer front. This process is called image warping.

A database of warped images can thereby be generated, and the process of estimating current articulation angle becomes an image 'look-up' problem. A suitable image similarity metric can be used to compare the current observed image with the images stored in the database, to determine the closest match and hence estimate Γ.

This method was first proposed by Harris in [73] but was not fully developed. In this work the method was further developed and refined, and an Unscented Kalman Filter (UKF) was added to smooth the noisy signal output of the template-matching algorithm.

This section begins with details of the image warping and image matching processes. The overall template-matching method is then summarised followed by brief discussions of incremental motion behaviour, large articulation angles and edge detection.

3.2.1 Image warping

Using the pin-hole imaging model it is possible to derive a relationship between the images of a planar object from two different viewpoints, provided the relationship between the two viewpoints is known. In this case this is the warping of a datum image, taken of the trailer at zero articulation, to yield an image of the trailer at non-zero articulation. It is assumed that a datum image of the front of the trailer can be obtained while the vehicle is driven forwards in a straight line.

Consider the rotation of the trailer relative to the camera as illustrated in Figure 3.3a. Identical relative motion between the trailer and the camera can be achieved through a virtual opposite rotation of the camera about the hitch point with an arc of radius $(h+d)$ as shown in Figure 3.3b. If it is assumed that the world co-ordinate system is aligned with the camera co-ordinate system at $\Gamma = 0^\circ$, the rotation and translation of the camera can be described by the R and $\mathbf{T}$ terms in Equation 2.8 respectively. These may be described in terms of Γ, h and d as follows:

$$\mathrm{R} = \begin{bmatrix} \cos\Gamma & 0 & \sin\Gamma \\ 0 & 1 & 0 \\ -\sin\Gamma & 0 & \cos\Gamma \end{bmatrix} \tag{3.14}$$

$$\mathbf{T} = \begin{bmatrix} (h+d)\sin\Gamma \\ 0 \\ (h+d)(1-\cos\Gamma) \end{bmatrix} \tag{3.15}$$

The relationship between the pixel co-ordinates of two views of a plane is given by

$$\tilde{\mathbf{w}} = \mathrm{P}\tilde{\mathbf{w}}_0 \tag{3.16}$$

where P is the 'planar homography' matrix between the two images and the subscript '0' indicates the image of the datum position ($\Gamma = 0^\circ$). P is defined as [168]

$$\mathrm{P} = \mathrm{K}\left[\xi\left(\mathrm{R} + \frac{1}{d}\mathbf{T}\mathbf{N}^\mathsf{T}\right)\right]\mathrm{K}^{-1} \tag{3.17}$$

where $\mathbf{N}$ is the normal to the observed plane at $\Gamma = 0^\circ$ (see Figure 3.3b), and ξ is the arbitrary scale factor due to the scale ambiguity in the $(\mathbf{T}/d)$ term. In this case,

$$\mathbf{N} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\mathsf{T} \tag{3.18}$$

Figure 3.4 shows the image of the trailer front at $\Gamma = 0^\circ$ (the datum image) and the effect of the warping process through the transformation, P. Before the warping step, the background must be cropped from the datum image to leave only the trailer face (also shown in Figure 3.4). This was done manually here, however in reality 'optical flow'[2] could be used to differentiate the trailer from the background during the same straight line manoeuvre used to obtain the datum image. This was not explored in this work.

The transformation P is a function of Γ, and so by varying Γ through the anticipated articulation range a database of warped images can be generated. A suitable increment may be chosen between warped images, which ultimately limits the precision of the look-up process. To achieve a resolution of 0.1° would require an increment of 0.1°. Smaller increments would unnecessarily increase the computation time required for creating the database.

The image warping process converts each pixel location (and associated intensity) from image co-ordinates to world coordinates, and then to image coordinates at a new view point. In reality integer pixel locations in the original image will not necessarily translate into integer pixel locations in the warped image, and so interpolation is used in many implementations of this process (such as in the VLFeat [169] and OpenCV [164] software libraries). Bilinear interpolation was used in this work.

It is worth noting at this point that the image of the trailer face at a given angle is dependent on the geometry of the trailer: the location of the hitch point relative to the camera and the front overhang of the trailer. These will affect the translational component of the relative motion between camera and trailer, which is fundamental to the image warping

[2]Optical flow is the apparent motion of image features or pixels between subsequent frames in an image sequence as the camera moves relative to the environment, or vice versa. In the case of the straight line manoeuvre, pixel motion would be detected in the background as pixels of the road surface, sky and roadside move towards the vanishing point in the image, but very little motion would be observed in the region of the trailer face as it is approximately stationary relative to the camera. This could be used to segment the datum image into trailer and background regions.

equations. Harris neglects this component and assumes that the variations in trailer geometry can be accounted for by incorporating a scale dimension only. Although this appears to work for the single case presented, the properties of perspective projection dictate that this was not an adequate solution.

3.2.2 Image matching

Once the warped image database has been generated and stored offline, the primary task of the algorithm is to compare each observed image with those in the image database and determine which database image provides the best match. The value of Γ at which the best matching database image is generated is then the estimated value of Γ. The entire database need not be searched, but rather a narrow search range based on the previous estimate may be used to reduce computation.

To determine the best match, a suitable similarity measure must be used. In computer vision, template-matching is the task of locating a known (usually smaller) template image within another image, indicating the existence of a particular object in that image. This is usually achieved through the combination of a sliding window search procedure and some image similarity metric. The image similarity metric gives a measure of the quality of the match at each location of overlap between template and image, so that the the location giving rise to the highest similarity determines the location of the template match within the search image.

For images of similar size, the magnitude of the similarity metric at the location of the template match may be compared between multiple images to determine which two images are the most similar. This is the concept used here in finding the best-matching database image.

Image similarity metrics

Template-matching was introduced in Section 1.7.4. This section focusses specifically on the different methods for evaluating the 'quality' of a successful template match. For the current application of template-matching, image similarity is the primary task and template match location is of secondary importance (though still useful). An extensive survey of image similarity methods up until 2003 is presented in [138] with relatively few additions to the field since then. A summary of these pre-2003 methods is discussed here, along with more recent developments.

In 'dense' template-matching, similarity can be determined through either a sum of differences technique or cross-correlation. Additional processing steps such as normalising

can also be included to reduce sensitivity to brightness, contrast and noise. These types of methods are often computationally demanding, and so not often suited to real time applications.

However, the computation of cross-correlation methods can be accelerated through calculation in the Fourier domain using the convolution theorem, such as in Lewis' 'fast normalised cross correlation' method [139]. Speed improvement in intensity-methods may also be achieved through the use of image pyramids—a sequence of down-sampled versions of the same image. This allows for an initial computationally efficient search in a low resolution version of the image in order to identify a region of interest in which to limit subsequent higher resolution (and more computationally expensive) searches. This is known as a coarse-to-fine strategy.

Edge detection (*e.g.* Canny edge detection [144]) can also be carried out prior to the correlation calculation. This can reduce the search region and improve precision and robustness to variations in brightness and contrast. It also enables a number of binary correlation techniques to be used for the comparison step (*e.g.* [170, 171]), and possibly contour detection and matching methods (*e.g.* [172–174]). Contour matching raises the possibility of identifying and matching geometric attributes of the trailer such as its outline.

Other 'dense' template-matching methods can be grouped together as information-based methods, which use distributed trends in the image data (*i.e.* pixel intensities or RGB values) for comparison purposes and do not necessarily consider the spatial structure of images. Methods include mutual information or entropy [140], histogram matching [141] and the 'perceptual hash' [175]. Dense methods in their basic form are not usually robust to large changes in scale, rotation, or out-of-plane warping, though some examples of these have been presented [138].

Feature-based methods involve the extraction of feature points, the subsequent matching of these features between the template and search image, and a measure of the similarity of the group of matched features. Feature types and common methods for feature detection and description were described in Chapter 2. The similarity of feature groups between images is often computed through spatial structure, using for example the Hausdorff distance metric.

Two recent developments in feature-based template-matching techniques include [170], a computationally efficient method using HoG features and Hamming distance which is currently not scale or rotation invariant, and the 'Best-Buddies Similarity' method [176], which is robust to transformations as well as background clutter and occlusion.

For many applications, feature-based methods are useful in that they are discriminative and invariant to brightness, orientation and scale. However, the assumption of available and distinct features is one disadvantage of feature-based methods, and intensity-based methods

have been shown to perform better in images where visual detail is finer in scale. This may be the case for HGV trailers, unless commercial branding or other visual features are visible.

For the current application, the primary criteria in choosing an image comparison method are a *lack* of invariance to transformations, robustness to a possible lack of distinct features, low computational demand and robustness to variations in brightness and contrast. It was concluded that normalised cross-correlation, specifically the computationally efficient implementation by Lewis [139], would be used. This offers no invariance to transformations, is not feature-based and so is potentially more robust to low-detail visual texture, and is robust to changes in brightness and contrast. The details of this implementation will now be described.

Fast normalised cross-correlation

With reference to Figure 3.5, $\gamma(u,v)$ is the 2-dimensional normalised cross-correlation (NCC) calculated at each point over which the template image, I_t, is 'swept' over the search image, I. The basic form of NCC is given by [139]:

$$\gamma(u,v) = \frac{\sum_{x,y}\left[\mathrm{I}(x,y) - \bar{\mathrm{I}}_{u,v}\right]\left[\mathrm{I}_t(x-u,y-v) - \bar{\mathrm{I}}_t\right]}{\sqrt{\sum_{x,y}\left[\mathrm{I}(x,y) - \bar{\mathrm{I}}_{u,v}\right]^2 \sum_{x,y}\left[\mathrm{I}_t(x-u,y-v) - \bar{\mathrm{I}}_t\right]^2}} \tag{3.19}$$

where:

- $\bar{\mathrm{I}}_t$ is the mean of the template,
- $\bar{\mathrm{I}}_{u,v}$ is the mean of $\mathrm{I}(x,y)$ overlapped by the template I_t at map coordinate (u,v), and
- $\sum_{x,y}$ is the sum over vertical and horizontal pixel coordinates in the region of overlap.

If I is an $n_u \times n_v$ image and I_t is an $m_u \times m_v$ template, γ will be a matrix of size $(n_u + m_u - 1) \times (n_v + m_v - 1)$, representing every possible pixel overlap. The extent of the overlap will range from $x \in 1$ and $y \in 1$ (only one pixel corner overlaps, as shown in top left and bottom right of Figure 3.5) to $x \in 1,\ldots,m_u$ and $y \in 1,\ldots,m_v$ when the full template overlaps the image, assuming the template to be always smaller than the search image. The warped image is taken to be the template and the observed image is the search image.

As Lewis noted [139], without normalisation NCC is a simple convolution of two signals, which can be computed efficiently by multiplication in the frequency domain in accordance with the convolution theorem [177]. With normalisation, the addition of the $\sum_{x,y}\left[\mathrm{I}(x,y) - \bar{\mathrm{I}}_{u,v}\right]^2$ term in the denominator prevents this computation in the traditional sense. However, the numerator of Equation 3.19 may still be computed in the frequency

domain as follows, where $\mathcal{F}()$ is the discrete Fourier Transform operator:

$$\overset{num}{\gamma}(u,v) = \mathcal{F}^{-1}\left\{\mathcal{F}\left(\mathrm{I}'\right)\mathcal{F}^{*}\left(\mathrm{I}'_{t}\right)\right\} \tag{3.20}$$

where $\mathrm{I}'(x,y) = \mathrm{I}(x,y) - \bar{\mathrm{I}}_{u,v}$ and $\mathrm{I}'_{t}(x,y) = \mathrm{I}_{t}(x,y) - \bar{\mathrm{I}}_{t}$.

Calculation of the denominator remains a computationally expensive task and cannot be achieved by frequency domain methods. Instead, Lewis proposed an algorithmic method to improve speed, the details of which are outlined in [139]. The net effect is a computational efficiency comparable to that of frequency domain computed cross-correlation, offering a significant improvement in the speed of the calculation, especially as the size of the template approaches that of the image.

The warped images of the trailer face are not rectangular, and hence background pixels were introduced to produce a rectangular image. These pixels (of arbitrary value) can lead to misleading results during the cross-correlation process and so must be removed. Figure 3.6 demonstrates the automatic cropping process adopted for this purpose. Each warped image was cropped to the width of the trailer face, and the height of the shortest side. Cropping the image by a few extra pixels at all sides was found to improve robustness, but too large a border was found to remove useful visual texture and reduce precision. An additional border of 20 pixels was found to be a good trade-off and was incorporated.

As the cropped template image was swept over the search image there was a peak in the correlation coefficient at the point where the image features overlapped the most. An example is shown in Figure 3.7a in which the peak is clear. The corresponding overlap of the template with the search image is shown in Figure 3.7b. This maximum correlation coefficient was used as the measure of image similarity between the two images.

To improve robustness and computation speed, the location of the maximum correlation coefficient in the correlation map from the previous time step was used to define a search window (or region of interest) in the current correlation map, within which to search for the current maximum. This is shown in Figure 3.7a. This effectively provided a limit on the distance in pixels which the trailer face was expected to move over a time step. This avoided erroneous matches between the template image and features in the background in the observed image by ignoring unreasonably large changes in match location, a problem observed in preliminary simulations.

For the low speed scenarios considered here, a window size of 60×40 pixels was found to be suitable. In future, a variable size window or more sophisticated tracking method may prove more flexible.

Edge detection

The use of edge detection before the NCC step was considered as a possible means to improve precision. In practice, this would be applied to all database images, and to each observed image as it is captured. The cross correlation step would then be carried out between two edge images instead of greyscale images. Edge detection can be performed by blurring an image with a Gaussian filter and then determining the contours of maximum pixel intensity gradients, producing a new binary image with line edges plotted along these contours. The scale at which edges are detected can be adjusted by adjusting the size of the Gaussian filter. The effect of the edge detection is shown in Figure 3.8 for the observed image (cropped).

The effectiveness of this approach was explored with a sample set of images from the CAD model. An observed image at 30° was compared with warped template images in a $\pm 1°$ range at 0.1° increments. 'Canny' edge detection was used [144] with lower and upper intensity thresholds of 30 and 90 respectively (change in intensity per pixel).

The results of the correlation analysis with and without edge detection are shown in Figures 3.9a and 3.9b respectively. As expected, the peak maximum correlation coefficient is at the correct template angle of 30°. However, the behaviour for the two cases is quite different as templates away from the correct value are assessed.

Also evident are oscillations in NCC with warp angle, resulting in localised 'peaks'. This will be henceforth denoted as 'aliasing', as a result of the integer values of pixel locations[3]. To demonstrate how this works in practice: if a single-pixel-width vertical line of intensity 1 warps to a mid-pixel location in the new image, the result will be a two-pixel-width vertical line of intensity 0.5. There will then be a slightly higher warp angle at which the vertical line will be reproduced again at intensity 1, shifted one pixel laterally.

If another image of a vertical line of intensity 1 was cross-correlated with all these warped images generated within a small range of warp angles (such that the width of the line itself does not vary significantly), then the resultant maximum cross correlation will exhibit multiple local maxima and minima as the line correlates at integer pixel locations and mid-integer pixel locations respectively. The width of the peaks correspond to the expected pixel widths at the respective articulation angles.

With edge detection the main peak is very pronounced, and maximum correlation coefficients drop rapidly away from the peak to values of around 0.4–0.6. Without edge detection, the drop is more gradual and smooth, and maximum correlation coefficients do not drop

[3]This definition of aliasing is not completely in line with the definition used more generally in signal processing and image processing, however they are related in that that they are both caused by a lack of sampling resolution, resulting in artefacts in the measured signals.

below 0.9. Edge detection therefore seems likely to offer superior precision, but there is an important trade-off in the presence of disturbances.

Consider Figures 3.9c and 3.9d—the same tests except with a larger window of 5°. Assume that some disturbance has resulted in a previous estimate of $\Gamma = 32°$ (shown) which is now the centre of the current search range, and that the current search range is 2° ($\pm 1°$). In the case of normal NCC, the smooth nature of the curve in this region is such that the estimate will quickly move towards the correct peak near 30°, over the next 2 or 3 time steps. In the edge detection case, the estimate is much more likely to settle at a local peak and won't be directed towards the correct peak over subsequent time steps. Further, additional disturbances may perpetuate the problem, resulting in an unstable drift towards large errors.

Consequently, edge detection was not carried forward as it seemed likely to result in robustness problems when implemented under realistic conditions.

3.2.3 Summary of the template-matching method

Methods for generating a database of warped images, and for matching an observed image with images in the database using NCC have been presented. The overall methodology is summarised in Figure 3.10. A hypothetical controller is shown to indicate how the system might be incorporated in a real-time control system such as reversing control or jackknife prevention.

In preliminary simulations, it was found that down-sampling the observed and template images by one pyramid level (effectively halving the resolution) had only a small effect on the accuracy of the system while having a significant positive impact on processing speed. Furthermore, using a search increment of 0.2° as opposed to 0.1° was also shown to have only a minor effect on performance while improving processing speed. Both these modifications were incorporated in the final implementation of the algorithm to improve frame rate, as this was otherwise impractically slow.

The template-matching algorithm was implemented in MATLAB [166] on a 64-bit Windows system with a 6-core 3.2 GHz Core i7 processor (see Appendix A) for details. The 'vl_imwbackward' VLFeat [169] function was used for image warping (with bilinear interpolation). The 'normxcorr2' MATLAB function, incorporating the acceleration methods of Lewis [139], was used for image matching. OpenCV [164] was used for its general image handling functionality.

640×480 resolution greyscale images were used. A search range of $\pm 1°$ and an increment size of 0.2° were found to be appropriate for the database look-up process. An additional crop border of 20 pixels was used for warped images. The region of interest

window size was kept at 60×40 pixels for locating the maximum correlation coefficient in each correlation map.

3.2.4 Incremental motion behaviour

It is useful to study the relationship between the rotational motion of the trailer and the resultant lateral movement observed in pixel co-ordinates. Consider Figure 3.11 which shows the trailer at angle Γ, and rotated by a small increment $\Delta\Gamma$ from this position. For the centre of the trailer face, the 'observed' lateral displacement is denoted e, where 'observed' means incorporating the perspective projection process, as if the imaging plane were located at the centre of the trailer face.

An expression for e as a function of Γ and $\Delta\Gamma$ may be found by trigonometry, using the accompanying geometric detail shown in Figure 3.11. Lengths and angles which can be calculated by simple trigonometric relationships are already indicated in the diagram. Starting in the right triangle A, σ_3 can be found as follows:

$$\sigma_3 = \tan^{-1}\left(\frac{h\sin\Gamma}{h+d-h\cos\Gamma}\right) \tag{3.21}$$

and then, using the sum of angles at an intersection, σ_1 is given by:

$$\sigma_1 = 90^\circ + \sigma_3 + \Gamma \tag{3.22}$$

By Pythagoras' theorem in triangle A, L_1 can be found:

$$L_1^2 = (h\sin\Gamma)^2 + (h+d-h\cos\Gamma)^2 \tag{3.23}$$

followed by L_2, by the cosine rule in the combined triangle BC:

$$L_2 = \sqrt{L_1^2 + h^2\tan^2\Delta\Gamma - 2L_1 h\tan\Delta\Gamma\cos\Gamma} \tag{3.24}$$

Finally, by the sine rule in triangle BC, σ_4 is given by the expression:

$$\sigma_4 = \sin^{-1}\left(\frac{h}{L_2}\tan\Delta\Gamma\sin\sigma_1\right) \tag{3.25}$$

and this enables σ_2 to be calculated by the internal angles of the triangle:

$$\sigma_2 = 180^\circ - \sigma_1 - \sigma_4 \tag{3.26}$$

Having found an expressions for σ_2, e can be found using the sine rule in triangle C, which simplifies to give the following equation:

$$e = -\frac{h \sin \sigma_2 \tan \Delta\Gamma}{\sin(\Gamma + \sigma_2)} \tag{3.27}$$

Given e, the lateral displacement in world co-ordinates, the resultant displacement in the image plane is denoted e'. Given that $k_u = 1$ for the CAD model (see Section 2.2), e' is equivalent to a lateral pixel displacement of Δu. Variables e' and Δu are related to e by

$$\Delta u(\Gamma, \Delta\Gamma) = e' = \left(\frac{f}{Z_{c,A}}\right) e \tag{3.28}$$

where $Z_{c,A}$ is the co-ordinate of the centre of the trailer face and is found to be

$$Z_c = d + h(1 - \cos\Gamma) \tag{3.29}$$

The above equations may be used to examine the behaviour of Δu as Γ and $\Delta\Gamma$ are varied, where $\Delta\Gamma$ may be interpreted as the increment size of the warped image database (provided it is small). Figure 3.12 shows the variation in Δu with Γ and $\Delta\Gamma$ for the CAD scenario. This was generated for selected values of $\Delta\Gamma$ $(0.05^\circ, 0.1^\circ, 0.15^\circ, 0.2^\circ)$ and for the entire range of Γ for the case of VGA resolution images (640×480). The figure will scale linearly with a variation in resolution.

It is clear from the plots that $\Delta u = 0$ for a particular magnitude of Γ which is denoted Γ_{lim} ($\approx 70^\circ$). This represents the point where the trailer face normal is perpendicular to the camera line of sight (see Figure 3.11). At this point, incremental movement of the trailer face is parallel to this ray and produces no lateral displacement in the image plane. Beyond this point the face is no longer visible. (Figure 3.12 shows the translation of the face in the opposite direction beyond this point, as the trailer continues to rotate.) From Figure 3.11, this is the point where $\sigma_2 = 0^\circ$ and so $e = e' = \Delta u = 0$. This point, Γ_{lim}, is a function of h and d only and can be calculated according to:

$$\Gamma_{lim} = \cos^{-1}\left(\frac{h}{h+d}\right) \tag{3.30}$$

Knowledge of this point is useful because it describes the limit beyond which the front face of the trailer can no longer be used for template-matching or other algorithms. For the simulated values of h (1.2 m) and d (2.3 m) this angle is 69.95°.

3.2.5 Large articulation angles

To estimate angles beyond Γ_{lim}, the trailer sides would need to be incorporated into the template-matching scheme. In Harris' results [73] it was clear that large errors can result at angles close to this point, especially when using the homography decomposition method.

Consider again the incremental motion behaviour of Figure 3.11. Points D and E can be taken to be the 'centres' of the trailer side faces, and for the moment it is assumed that datum images of the trailer sides are available. Using the same geometrical approach as above, but with h replaced by g and d replaced by $(d+h)-g$ (see Figure 3.11), the incremental motion behaviour for the trailer sides may be calculated.

Results for both the trailer front and sides for 0.1° articulation angle increments are shown in Figure 3.13. The limit angle for the trailer sides is $|\Gamma_{lim}| = 20.6^\circ$; $|\Gamma_{lim}| = 69.95^\circ$ for the trailer front as before. There are therefore two ranges between -69.95° and -20.6° and between 20.6° and 69.95° in which both the front face and a side face are visible.

A higher value of pixel movement per increment translates into a better sampling ratio between pixel space and angular trailer motion which should provide more robust measurement. Therefore the ideal point at which to transition from template-matching based on the front face to the visible side face is the point at which these curves cross. These transition points are denoted Γ_{trans} and in this case $\Gamma_{trans} = \pm 44.8^\circ$.

The template-matching algorithm could be extended to accommodate this procedure. Some improvement in accuracy might be expected given that narrow template images of the trailer front near Γ_{lim} could be avoided, and side datums could be used at that point instead (or front and side results could be averaged in this region).

Incorporating side images in such a way was explored by Harris [73], however results demonstrated very large errors in the region of transition. Although it may be possible to improve the transition method from that used by Harris, the requirement for side datum images from a practical point of view is limiting, and would require additional calibration steps or additional information about the trailer. As such only the front images were used, and TM was not assessed beyond $\pm\Gamma_{lim}$.

For this reason investigation of the template-matching algorithm was restricted to the use of the front face, and hence it was limited by geometry to operate within Γ_{lim}. For the current simulation geometry this still accommodates a large range of close to $70^\circ \leq \Gamma_{lim} \leq 70^\circ$. Other geometries may be further restricted, but it is likely that the range will be sufficient for most typical driving manoeuvres. Exceptions might include reversing around a tight corner.

3.3 Unscented Kalman Filter

Preliminary simulations of the template-matching method exhibited noise in the output Γ signal. An Unscented Kalman Filter (UKF) [178] was incorporated to smooth these noisy measurements. In this section details of the filter and the underlying vehicle model are described. The combined template-matching and UKF algorithm will be referred to as 'TM+UKF', to differentiate it from the unfiltered TM algorithm.

3.3.1 Kinematic vehicle model

A non-linear kinematic vehicle model was used so as to have the simplest model to capture the low-speed yaw motion of the vehicle accurately. This is relevant in a commercial context in which vehicle parameters may have to be estimated. The model incorporates non-linear articulation angle behaviour at higher angles.

The 3-degree-of-freedom kinematic 'bicycle model' used is illustrated in Figure 3.14. As before, c is the hitch offset, Ω is yaw rate, and U and V are the longitudinal and lateral velocities at the rear axle respectively. The wheelbase is l. Tractor and trailer parameters are subscripted 1 and 2 respectively. The steer angle is δ and the articulation angle is Γ.

By kinematic arguments, we note that:

$$\Omega_1 = (U_1/l_1) \cdot \tan\delta \tag{3.31}$$

$$\Omega_2 = \Omega_1 + \dot{\Gamma} \tag{3.32}$$

$$V_2 = (V_1 + c_1\Omega_1) \cdot \cos\Gamma - U_1 \sin\Gamma - \Omega_2 l_2 \tag{3.33}$$

The kinematic assumption, which is reasonable at low speeds, requires that $V_1 = V_2 = 0$. Assuming U_1 to be constant, combining Equations 3.31–3.33 yields the following non-linear state equation:

$$\dot{\Gamma} = -U_1 \left(-\frac{c_1}{l_1 l_2} \tan\delta \cos\Gamma + \frac{1}{l_2} \sin\Gamma + \frac{1}{l_1} \tan\delta \right) \tag{3.34}$$

To account for tyre scrub effects of multiple unsteered axles in an axle group, the equivalent wheelbases concept was used [179]. The wheelbases l_1 and l_2 may be redefined as follows:

$$l_{eq} = l + \frac{TF}{l} + \frac{TF}{l} \left(\frac{C_r}{C_f} \right) \tag{3.35}$$

where C_f and C_r are the sums of front and rear tyre cornering stiffnesses (assumed linear)[4]. TF is the 'tandem factor' and is defined as:

$$TF = \frac{\sum_{p=1}^{n_a} \rho_p}{n_a} \tag{3.36}$$

where ρ is the axle spacing measured from the geometric centre of the axle group and n_a is the number of axles in the group.

3.3.2 Filter description

The Unscented Kalman Filter (UKF) [180, 181] is a non-linear filter which, unlike the Extended Kalman Filter (EKF), does not require the derivation of Jacobian matrices, analytically or otherwise. Rather, the non-linearities of the system model are accounted for through the use of carefully chosen 'sigma points' which accurately capture the mean and covariance of the random variable being propagated. The UKF is at least second-order accurate whereas the EKF is only first-order accurate, and yet it maintains a comparable computational efficiency to the EKF [181]. Full details of the UKF are given in Appendix B.

In the case of the kinematic bicycle model the state vector may be simply defined as $\mathbf{x} = \Gamma$. In discretised form, Equation 3.34 is the process model, $\mathbf{F}$. The measurement model, $\mathbf{H}$, is simply $\mathbf{y}_k = \Gamma_{TM}$ where Γ_{TM} is the noisy template-matching measurement output. Measurement and process noise were assumed to be zero-mean Gaussian with covariance W and Q respectively.

For non-constant speeds, U_1 may be assumed to be constant over a time step. Therefore, although Equation 3.34 assumes constant velocity, instantaneous values of U_1 were used in the UKF. The filter was implemented in MATLAB alongside the template-matching algorithm.

3.4 Parallel Tracking And Mapping method

The second image processing method explored for articulation angle measurement was Parallel Tracking and Mapping (PTAM). Unlike template-matching, this is a 3-D approach which uses feature tracking to estimate the 6-degree-of-freedom pose of a camera relative to an observed scene. An outline of the algorithm is presented here, along with details of the implementation of the algorithm for the current task.

[4]Tyre cornering stiffness is load-dependent, and so this introduces a sensitivity to loading conditions. For trailers, the third term in Equation 3.35 may be disregarded [47], but the issue remains for the truck. However, in Section 3.6.4, sensitivity to the equivalent wheelbase (and hence tyre stiffness) is shown to be negligible.

3.4.1 Introduction

The PTAM algorithm [123] was developed for augmented reality applications using hand-held cameras. It is a solution to the monocular SLAM problem (Simultaneous Localisation And Mapping[5]) and is an alternative to existing SLAM approaches such as EKF-SLAM and FastSLAM. In the context of mobile robotics and augmented reality, the SLAM problem is one of estimating the pose of a moving agent within a predominantly static scene ('localisation'), while creating a map of the scene ('mapping'). The two tasks are interdependent and cannot be performed entirely in isolation of each other.

The SLAM problem is not fundamentally dissimilar to the problem of trailer motion sensing. Here the goal is to determine the pose of the trailer relative to the tractor. The camera is static (relative to the tractor) and the scene (the trailer) is moving but not deforming. Provided the trailer occupies the majority of the camera field of view, the problem is interchangeable.

PTAM differentiates itself from typical SLAM approaches primarily through the parallel nature in which it performs the two tasks of mapping (generating a 3-D map of the observed surroundings) and tracking (locating the camera's pose within the surroundings). The two tasks are performed in parallel processing threads with only intermittent communication between the two [123].

Unlike the 'frame-by-frame' approach of typical SLAM algorithms, mapping is not updated at every frame but is limited only to certain 'keyframes' where sufficient camera translation has occurred for the mapping update to be meaningful. This allows for a computationally more expensive but more accurate method of mapping to be performed. In addition, by taking mapping out of the tracking thread, more computational time is made available for the tracking task, resulting in more accurate camera pose estimation [123].

Murray and Klein's [123] application of PTAM to augmented reality for hand-held cameras required the observed scene to contain a dominant planar surface. However, this was only a requirement of that particular application, so as to introduce ground-based virtual agents into the scene. The PTAM algorithm itself makes no assumptions regarding planar surfaces and is theoretically applicable to arbitrarily non-planar scenes.

This means that the algorithm would be applicable to theoretically all trailer shapes, including box-type trailers, tankers with domed fronts, trailers with front-mounted refrigeration units, and possibly even abnormal load trailers or car-carriers. As more of the trailer is viewed with increasing articulation angles, additional keypoints can be added to the three-dimensional map of the trailer, allowing for efficient recall when viewed again.

[5]The 'localisation' in SLAM and 'tracking' in PTAM can be taken to have the same meaning: determining the 3-D pose of the camera relative to the scene.

Loop closure (Section 1.7.2) is an important aspect of SLAM-type algorithms, especially in the context of mobile robotics. Loop closure is the ability to recognise a pre-visited scene map and hence reduce the pose and map errors which have accumulated since visiting that location previously. Although this remains a difficult problem, it is of little consequence to this application. The observed scene is unchanging, except when articulation angles are sufficiently large such that the front face of the trailer (i.e. that of a flat-fronted trailer) disappears from view and is replaced with a view of the side of the trailer. This is not deemed to be problematic as the original scene is not diverged from significantly.

3.4.2 Theory

A detailed description of PTAM's underlying theory can be found in [123]. A summary is given here.

PTAM consists of two parallel processing threads: a mapping thread and a tracking thread. Assuming a map of 3-D feature points has already been generated, the tracking thread is responsible for matching detecting features in the current frame with features observed in previous frames, and thereby updating the camera's pose in the current frame. Thus the motion of the camera is tracked in a known 3-D scene. Tracking is performed at frame rate (*i.e.* at each new frame).

Details of the tracking task may be summarised as follows:

1. FAST features [182] are detected in the current image at four image pyramids levels (*i.e.* four versions of the same image, each down-sampled by a factor of two relative to the preceding level). A small image patch around each feature point is also recorded, assuming the feature to be locally planar.

2. A 'prior' estimate of the camera's pose (position and orientation) is calculated based on its previous position using a decaying velocity motion model (similar to an alpha-beta filter [183]).

3. Known 3-D map points are projected into the current image frame. The pin-hole model (see Equation 2.8) is used and distortion is accounted for using the FOV-model [160] (Section 2.3). The planar feature patches are warped to accommodate the change in camera pose (similar to Equation 3.17 but using a simplified affine rather than full perspective model).

4. These projected points are matched to detected points in the current image using a fixed-range search around their predicted locations. An initial coarse search is done with only 50 features.

5. The camera pose is updated using the matches of the 50 points. A further update of the pose is performed using up to 1000 remaining points, searching over a smaller window.

The mapping thread is responsible for building the 3-D map of feature points relative to the global reference frame defined during initialisation. This does not run at frame rate (nor does it need to), but rather performs more intensive map updates when necessary, independently of the tracking thread. The mapping task may be summarised as follows:

1. The first map is built initially using a stereo technique. This is the stereo initialisation step. Two images of the same scene are manually captured with a moderate translation of the camera between them. These two images are the first two keyframes. Because the magnitude of this motion is not known exactly, assumptions are made and the map is only reproduced up to an arbitrary scale factor.

2. The map is continually expanded and refined with the addition of new keyframes. New keyframes are added when necessary, provided at least 20 frames have passed since the last keyframe and there is a minimum distance from other keyframes. This avoids the pose drift behaviour caused by image sequences from a stationary camera [123].

3. For every new keyframe, local bundle adjustment [124] is performed to iteratively optimise reprojection errors over all map points and keyframes.

4. When no new keyframes are being added, and the camera is viewing previously explored sections of the map, the mapping thread uses this 'free time' to refine the existing map. This is done using global bundle adjustment, which is too computationally expensive to be calculated in real-time.

The stereo initialisation step is important, as this defines the initial map upon which all updates are based. Features are detected in an initial image, making up the first keyframe, and these features are then continuously tracked. A small translation of the camera relative to the scene is then required before the second keyframe is taken, capturing the new locations of these features. For typical hand-held camera applications, Klein and Murray assume the size of this translation in world co-ordinates to be 10 cm in order to generate the initial map using the five-point algorithm [184]. As the exact magnitude is not actually known and can vary, the resultant Cartesian map is only accurate up to a scale factor. Rotations are independent of this assumption.

3.4.3 Modification and implementation

The C++ source code for PTAM is open source and required relatively minor modifications for application here. The algorithm provides two primary outputs: a 3-D map of feature point locations, and the camera pose in the form of a 3×1 translation vector and a 3×3 rotation matrix. The rotation matrix is relative to the dominant plane observed during initialisation (with arbitrary datum), and the translation vector represents the motion of the camera relative to its original location.

Of interest here is the rotation matrix and particularly the yaw angle of the camera relative to the observed scene (*i.e.* the trailer). A general 3×3 rotation matrix can be decomposed into a combination of sequential rotations about each axis, known as the Euler angles: roll (ϕ), yaw (ψ) and pitch (θ). This is summarised as follows:

$$\mathrm{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} = \mathrm{R}_z(\phi)\mathrm{R}_y(\psi)\mathrm{R}_x(\theta) \tag{3.37}$$

where

$$\mathrm{R}_z(\phi) = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathrm{R}_y(\psi) = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix}$$

$$\mathrm{R}_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

Euler angles may be extracted from the rotation matrix using a combination of the above definitions and logic to reject multiple solutions. The method of [185] was used for this purpose.

The above rotations are in the camera co-ordinate frame. The yaw angle may be taken to be the yaw angle of the trailer relative to the tractor provided pitch and roll variations are small. In the case of pitch and roll however, one must take care to note the effect of the transformation from camera to trailer co-ordinates. For example, assume that the trailer is pitched at 3° about its own lateral axis. In the camera reference frame this would register as a 3° pitch angle only when the articulation angle is zero. At an articulation angle of 90°, this would register as a 3° *roll* angle.

More generally, this can be accounted for as follows:

$$\begin{bmatrix} \phi \\ \psi \\ \theta \end{bmatrix}_{trailer} = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix} \begin{bmatrix} \phi \\ \psi \\ \theta \end{bmatrix}_{camera} \tag{3.38}$$

The existing motion filter of the PTAM algorithm (a decaying velocity model) was found to work acceptably well and there was no need to modify it or introduce an UKF as before.

The algorithm was implemented in C++ on a 64-bit Linux laptop computer with a dual-core 2.4 GHz Core i3 processor and 4GB of RAM (detailed specifications can be found in Appendix A).

3.5 Simulation scenarios

To generate simulation data, sinusoidal steer signals were input to the dynamic vehicle model to generate articulation angle responses. The articulation angle signals were then input to the CAD model to generate the accompanying image sequences for processing by the two algorithms. Speed was kept constant at 6 km/h. The steer input was also used as the input for the UKF, with added noise of standard deviation 0.1°. For the UKF, the measurement noise covariance, W, was determined experimentally to be $(0.48^\circ)^2$. The process noise, Q, was tuned to $(0.06^\circ)^2$ through trial and error on a few initial datasets.

Three tests were conducted:

1. A reference test with $\Gamma < \Gamma_{lim}$ ($< 69.95^\circ$), with no UKF, no down-sampling and using an image search increment of 0.1°. Although this is not the proposed configuration of the template-matching algorithm and is too slow for practical purposes, the results provide useful insights into its performance.

2. The reference test was repeated, with one stage of down-sampling and an image search range of 0.2° for the TM+UKF algorithm. This was also processed using the PTAM algorithm.

3. A test with large articulation angles (up to almost 90°) was conducted to assess PTAM's performance once the sides of the trailer become visible. This test was processed with PTAM only.

The stereo initialisation process required by the PTAM algorithm was carried out manually at the beginning of each test run. The first keyframe was obtained for an image close to $\Gamma = 0^\circ$ at the start of the test. Once the trailer began to rotate, the second keyframe was

captured after some lateral displacement of the features was evident (likely all in excess of 10 cm). No further manual input was required beyond this point.

In addition, a parametric study of the sensitivity of TM+UKF algorithm variations of $\pm 10\%$ and $\pm 20\%$ in the values of trailer front overhang (h), hitch location (d, c_1) and trailer wheelbase (l_2) was carried out.

3.6 Simulation results and discussion

3.6.1 50° reference simulation (TM only)

A time history plot of the TM results for the reference simulation is shown in Figure 3.15 with the accompanying error signal. Error variation as a function of articulation angle is shown in Figure 3.16. The RMS (Root Mean Squared) and maximum errors were 0.26° and 0.69° respectively, and the average frame rate was 1.3 fps. The standard deviation of errors was 0.2°. RMS errors were within the 0.4° target, and maximum errors just above the target of 0.6°. Frame rate was expected to be low giving the lack of down-sampling and the 0.1° search increment.

The error at the first data point ($\Gamma = 0°$) was 0° as expected. Here the observed image was identical to the 0° warped image with the exception of cropping. For the remainder of the test however, the signal is seen to be noisy. Given the idealised nature of the simulations with only a yaw degree of freedom and no image distortion, the primary cause of errors must have been either due to the image warping or image matching processes. The source of this was found to be aliasing, as described in Section 3.2.2.

This effect is best exhibited by considering plots of the variation of maximum correlation coefficient within the search range at various points throughout the simulation. Figures 3.17a and 3.17b show the aliasing behaviour at low and high values of Γ respectively. The peaks and troughs are clear. At $\Gamma \approx 3.5°$ the spacing between peaks is approximately 5 increments, or 0.5°. At $\Gamma \approx 41.5°$ the spacing is approximately 9 increments, or 0.9°. Both these results agree with the incremental motion behaviour exhibited earlier in Figure 3.12.

Figure 3.18a shows how a good result (small error) was obtained when the true value of Γ was near a peak. Similarly, Figure 3.18b shows how a poor result was obtained when the true value of Γ was near a trough. Further, when the true value of Γ lies near a trough, the two peaks either side of it are often similar in magnitude, resulting in large 'jumps' in error (equal to the peak spacing in magnitude) as the estimate jumps from one peak to the other on sequential time steps. This is demonstrated in Figure 3.19 for $t \approx 24$ s. This sharp jump in

error is clear in Figure 3.15, and was repeated cyclically at the same value of Γ. Figure 3.16 further demonstrates this at $\Gamma \approx 38^\circ$.

The variation of error with articulation angle exhibits some clear striations (Figure 3.16). These appear to have been a result of the aliasing process. As new observed images are compared to the database, the distribution of the peaks and troughs changes. It is therefore possible that a match follows one peak as the peak moves, before jumping to an adjacent peak. For $-\Gamma$, the peak movement is exhibited in the individual striations, and the jumps are exhibited by the vertical jumps in error between striations. The behaviour appears slightly different for $+\Gamma$, but the magnitude of the vertical jumps is still coherent with the aliasing explanation. The nature of the horizontal spacing between striations is unclear. The asymmetry about $\Gamma = 0^\circ$ is likely a result of imperfect camera misalignment in the CAD model (described in Section 3.1.2).

In the absence of any external disturbances therefore, the magnitude of errors exhibited are constrained to the magnitude of the aliasing effect, which in turn is a function of the system geometry and image resolution. For the current geometry the maximum peak spacing is 0.45° which correlates well with the observed noise in the error signal (a standard deviation of 0.48°). It can therefore be concluded that in the ideal case, the accuracy of the template-matching algorithm is limited by aliasing.

3.6.2 50° simulation (TM+UKF, PTAM)

Results for the 50° simulation are shown in Figures 3.20 and 3.21. Template-matching results with and without the UKF (TM, TM+UKF) and PTAM results are shown. Figure 3.21 shows PTAM and unfiltered TM results only. For the TM+UKF case, RMS and maximum errors were 0.30° and 0.73° respectively, down from 0.49° and 1.65° without the filter. Average frame rate was 7.3 fps, meaning that measurements could be provided to an external control processor at 7.3 Hz, at a time delay of 0.14 s.

Errors in the unfiltered signal approximately doubled with the introduction of down-sampling and a larger search increment as expected, but with the benefit of a near six-fold increase in frame rate. The UKF effectively corrected for this loss in precision, yielding a good combination of accuracy and frame rate.

The PTAM results exhibit very low noise (due to its inherent filtering), and yield RMS and maximum errors of 0.60° and 1.14° respectively. These are comparable to the unfiltered TM results, but approximately twice those of the TM+UKF case. The error varies consistently as a function of articulation angle, suggesting some inaccuracy in an underlying calibration parameter or the initialisation process. Frame rate was a consistent 20 fps, significantly better than the TM+UKF algorithm.

The variation of TM errors as a function of articulation angle exhibits similar striations to before, but with vertical jumps equivalent to the new aliasing characteristics. These are clearest for $-\Gamma$, with jumps now approximately 1° in size, doubled in accordance with the single-stage down-sampling. The change in search increment does not affect the aliasing behaviour. The horizontal spacing between striations appears to have increased too, suggesting that these are also a result of the aliasing process. The asymmetry is attributed to camera misalignment as before.

The variation of PTAM errors with Γ is smooth, varying in an approximately sinusoidal manner and with some hysteresis. To some extent, errors are expected to increase with increasing Γ, as PTAM's ability to accurately match features from the first keyframe (near $\Gamma = 0^\circ$) to the current oblique keyframe deteriorates. Small misalignments in the camera could have resulted in camera rotation measurements not being aligned with the plane of rotation of the trailer, which would introduce an approximately sinusoidal variation in errors, possibly explaining the shape of the error trend here. The hysteresis is likely a side-effect of the tracking algorithm.

3.6.3 90° simulation (PTAM only)

Results for the 90° test are shown in Figures 3.22 and 3.23. This test was only applicable to PTAM given that Γ exceeds Γ_{lim} (69.95°). RMS and maximum errors of 0.79° and 1.56° were obtained, marginally higher than the 50° test case. The error trend as a function of articulation angle (Figure 3.23) is comparable to the 50° case in shape and magnitude, and the same sources of error are suggested.

Keyframes from the trailer sides were detected in these tests unlike before, but these were obtained at a more oblique view to the keyframes from the trailer front captured during stereo initialisation. It is therefore possible that the error probability in these feature locations is higher, contributing to increased errors overall for the large angle simulations. Longer simulations could possibly show evidence of improved accuracy due to additional bundle adjustment within PTAM's mapping thread. Even so, performance is good and the algorithm is able to cope with high values of Γ beyond Γ_{lim}. This is a significant benefit over the TM+UKF algorithm.

Figure 3.24 shows additional details of the PTAM simulation. The initialisation process is shown in Figure 3.24a, where features are detected and tracked over a relatively small lateral translation of the trailer. The detected feature points on the front and sides of the trailer are shown in Figures 3.24b to 3.24d. Initially only features on the front are in the map; features on the trailer sides are added to the map as they become visible at higher values of Γ.

The 3-D map of feature points is shown in Figure 3.24e, along with relative locations of the camera at each keyframe. A top view is shown in Figure 3.24f, in which some outlying features are evident, especially in the features on the trailer sides. The errors in the estimated locations of these feature points will have contributed to the overall pose error as discussed previously. The worst outliers are on the side of the trailer which is viewed at the highest values of Γ. When these features were added, the front of the trailer (and hence the entire initial map) had disappeared from view, making the mapping less accurate in this region and hence reducing the accuracy of camera pose estimation over the entire map (*i.e.* all articulation angles).

A summary of performance in all tests is given in Table 3.2.

3.6.4 Parameter sensitivity

The 50° test was used as the basis for a sensitivity analysis of the TM+UKF algorithm. The effect of 10% and 20% variations in assumed values of trailer front overhang, hitch location, and trailer wheelbase on RMS and maximum errors was studied. These were deemed to be the most critical of the assumed parameters on which the TM+UKF algorithm is dependent. Front overhang forms the basis of the image warping process, hitch location also forms the basis of the image warping process and is also a parameter required by the UKF, and trailer wheelbase is only required by the UKF.

Results are shown in Figure 3.25. Sensitivity to front overhang and hitch location is significant as expected with up to 4-fold increases in errors. In all cases however the algorithm suffered no robustness issues. It should be noted that the scaled values of these variations are large: in this case a 20% variation in front overhang equates to 240 mm, and a 20% variation in hitch location equates to 460 mm. In all cases RMS errors did not exceed 2°.

It is likely that these values may be assumed known to within variations smaller than those considered here, especially in countries where vehicle dimensions are tightly controlled. In countries were vehicle dimensions are less tightly controlled, additional methods for estimating or refining these values from the measured image data may be required.

There is only a small sensitivity to trailer wheelbase, which suggests that the performance of the UKF may be fairly robust to vehicle parameter assumptions. Provided measurements of steer angle and speed are available with suitable accuracy, the filter should perform well at reducing noise and overall errors over a range of operating conditions.

Note, however, that steer angle sensing is not always available on tractor units, so this may require an additional transducer in some cases. A simpler alpha-beta type filter as used in PTAM may prove sufficient in future work, and would remove the need for steer angle sensing.

3.6.5 Performance benchmark

A comparison of performance with the published literature is shown in Figure 3.26. The algorithms are compared against simulation results only from [30, 52, 73–75]. Results from Bouteldja *et al.* [30] and Chu *et al.* [52] are based on state observers. Results from [30] are in fact for tractor yaw estimation, not articulation angle, but these are deemed to provide an acceptable comparison. Results from Harris [73] are based on two visual methods explored with a similar camera configuration. All of Harris's results required excessive computational effort with frame rates below 0.2 fps. Recall that the system of Fuchs' *et al.* [74, 75] requires special trailer markers.

The only simulation results which perform better in terms of maximum error are one of the tests of Chu *et al.* and the results of Fuchs *et al.* However, only articulation angles of up to 3° were assessed in [52], and [74, 75] requires impractical special markers. Only one of Harris's results is comparable in performance to the TM+UKF and PTAM methods, but this runs at very low frame rates. Overall both the TM+UKF and PTAM algorithms exhibit very favourable performance over a large range of articulation angles compared to methods in the published literature.

3.7 Conclusions

1. Two camera-based concepts for articulation angle estimation for articulated HGVs have been developed, using a tractor-based rear-facing camera:

   (a) a template-matching algorithm initially proposed in [73], with additional development and the addition of an Unscented Kalman Filter, and

   (b) the PTAM algorithm of [123], developed for augmented reality applications, with novel application here.

2. In CAD-based simulations, a MATLAB implementation of the template-matching algorithm exhibited RMS and maximum errors of 0.30° and 0.73° respectively at 7.3 fps. The algorithm is limited to planar trailer fronts and a maximum articulation angle (dependent on trailer geometry), and requires some knowledge of vehicle geometry to which its accuracy is sensitive. Precision was shown to be limited by aliasing. Processing speed could be improved through implementation in C++.

3. The PTAM algorithm (implemented in C++) exhibited RMS and maximum errors of 0.60–0.79° and 1.14–1.56° respectively at 20 fps. The algorithm makes no assumption

about the planarity of the trailer front, requires no knowledge of vehicle geometry, and operates throughout the full range of articulation angles.

4. In this idealised environment with no disturbances, TM+UKF comes closest to the target measurement accuracies, and being intensity-based it is potentially more robust to low visual texture trailers. However, PTAM offers significant benefits in terms of flexibility and computational efficiency, resulting in greater potential for commercial implementation.

5. Performance of both algorithms was shown to be favourable compared to the published state-of-the-art.

3.8 Tables and figures

Table 3.1 Simulation vehicle parameters

Parameter	Tractor (1)	Trailer (2)
Mass, m_1, m_2	7728 kg	8805 kg
Yaw moment of inertia, J_1, J_2	26478 kg·m^2	84000 kg·m^2
Wheelbase, l_1, l_2	3.7 m	9.7 m
Number of rear axles, n_a	1	1
CoG location rear of steer axle/hitch, a_1, a_2	1.051 m	3.0 m
Hitch offset ahead of tractor rear axle, c_1	0.775 m	-
Front overhang, h	-	1.2 m
Tyre cornering stiffness co-efficient, C_f, C_r, C_t	201 kN/rad	
Camera to trailer front face, d	2.3 m	

Table 3.2 Results summary: planar vehicle tests

		TM		TM+UKF		PTAM	
Test	Γ_{max} (°)	ε_{max} (°)	ε_{RMS} (°)	ε_{max} (°)	ε_{RMS} (°)	ε_{max} (°)	ε_{RMS} (°)
50° (ref.)	50	0.69	0.26	-	-	-	-
50°	50	1.65	0.49	0.73	0.30	1.14	0.60
90°	84	-	-	-	-	1.56	0.79

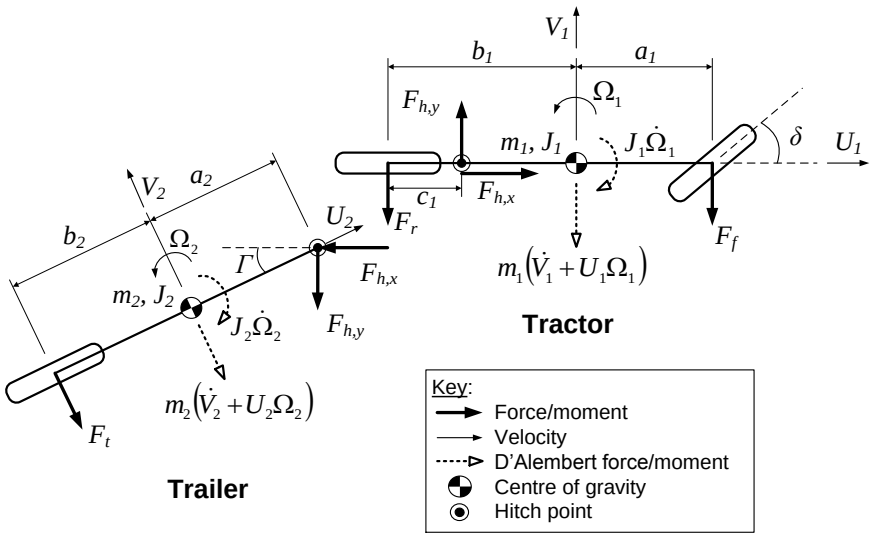


Fig. 3.1 3-degree-of-freedom dynamic bicycle model of a tractor semi-trailer

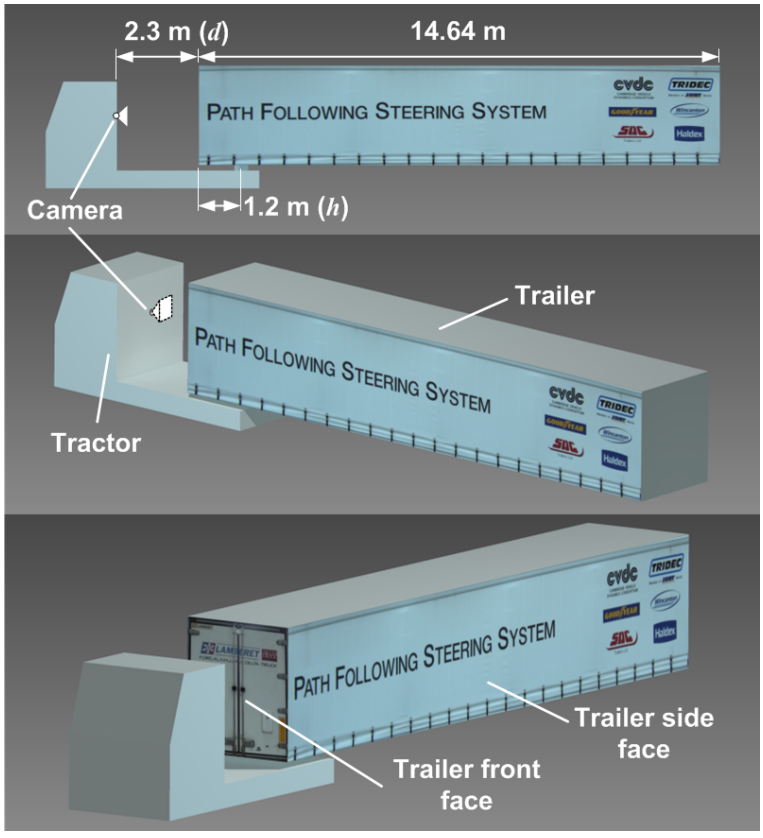


Fig. 3.2 CAD-based visual simulation environment

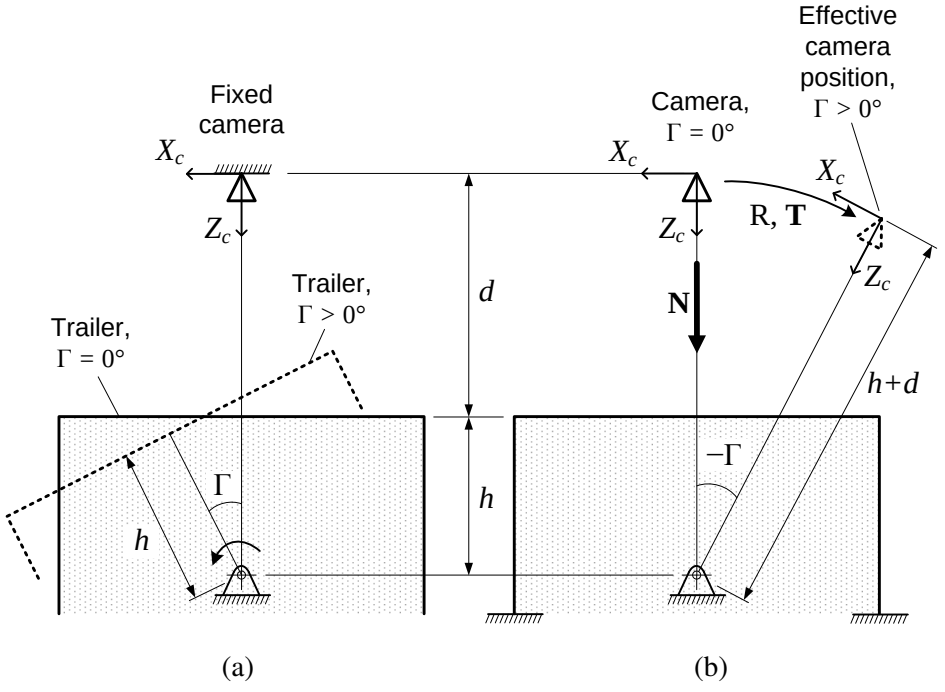


Fig. 3.3 Trailer articulation model: (a) actual trailer motion, (b) effective camera motion relative to trailer

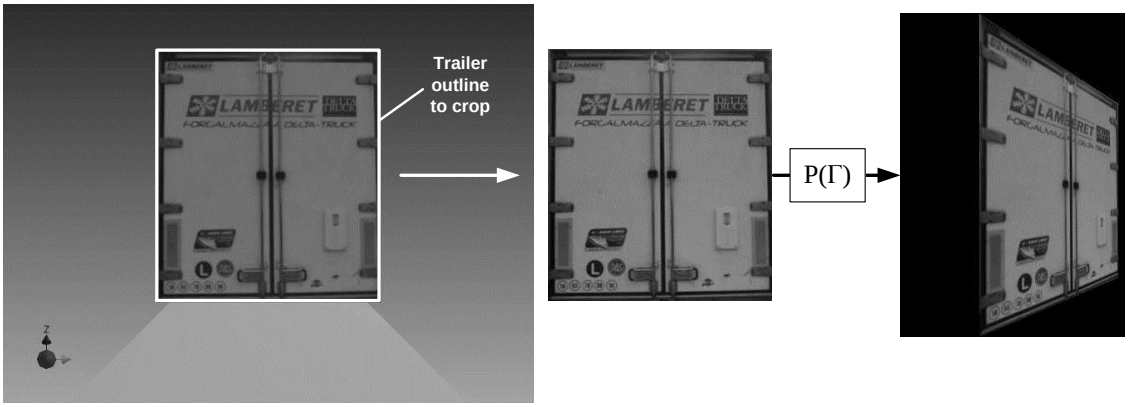


Fig. 3.4 Cropping the trailer face from the datum image, and warping it through transformation P ($\Gamma = 37°$)

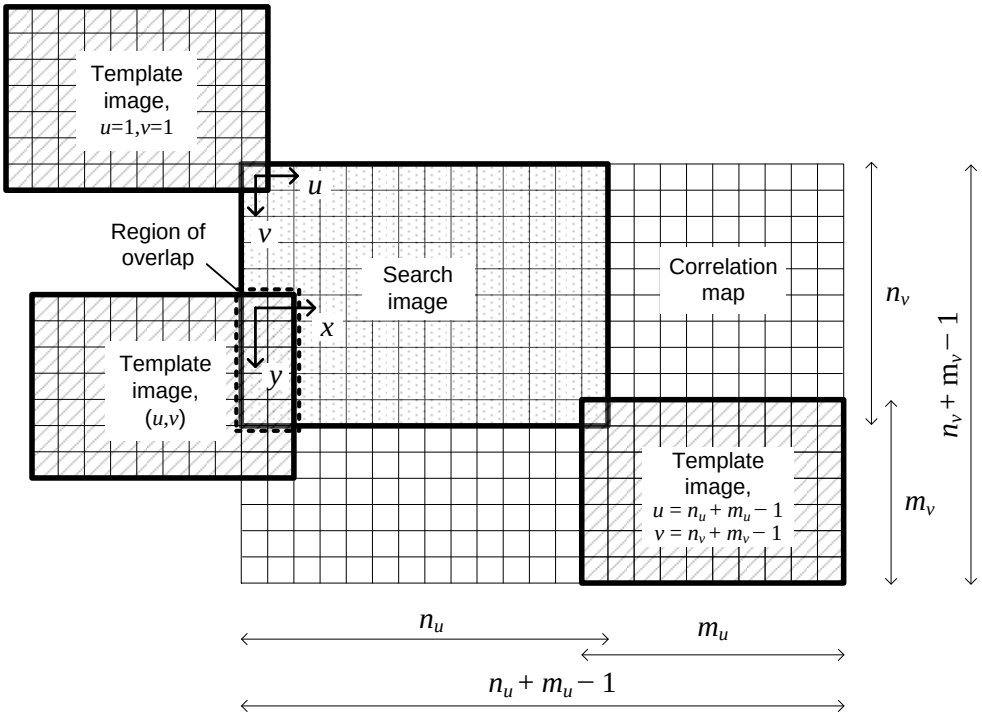


Fig. 3.5 The cross correlation process, sweeping the template image over the search image

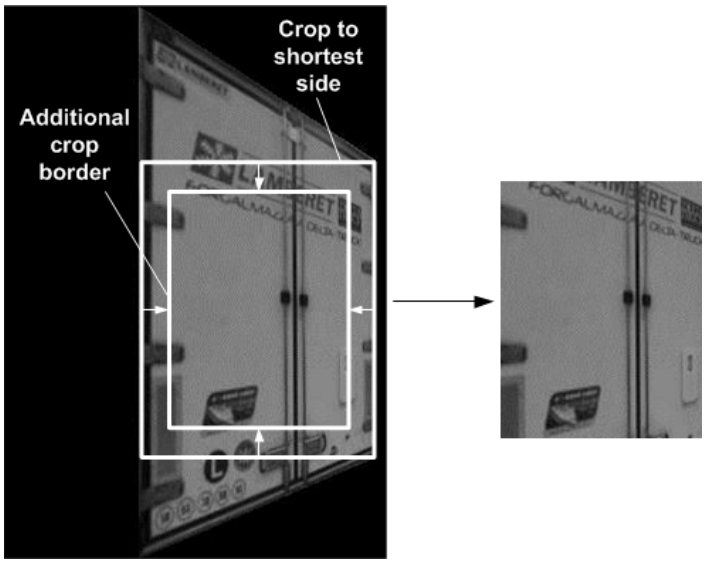
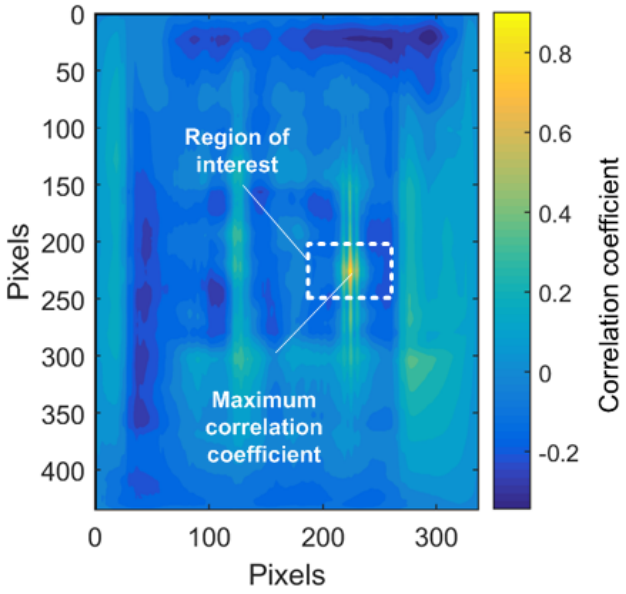
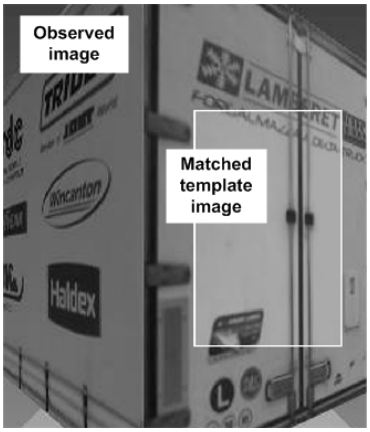


Fig. 3.6 Template image cropping ($\Gamma = 37°$)



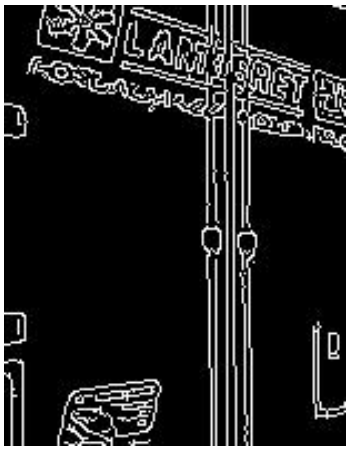
(a) Correlation map



(b) Template match location

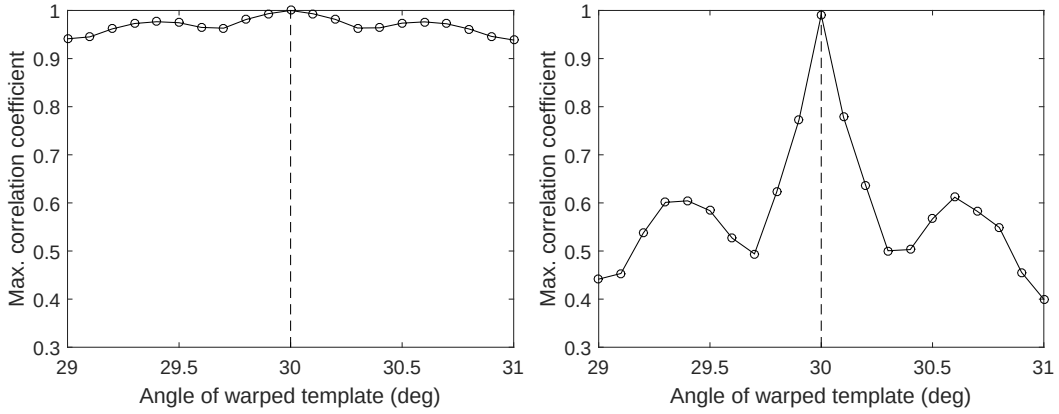
Fig. 3.7 Normalised cross-correlation between image and warped template ($\Gamma = 37°$)

(a) Greyscale image before edge detection



(b) Binary image after edge detection

Fig. 3.8 Canny edge detection on an example trailer image



(a) NCC, range = 2°

(b) NCC + edge detection, range = 2°

(c) NCC, range = 5°

(d) NCC + edge detection, range = 5°

Fig. 3.9 The effect of edge detection on the profile of maximum correlation coefficients in the search range. Correct match shown with dashed line

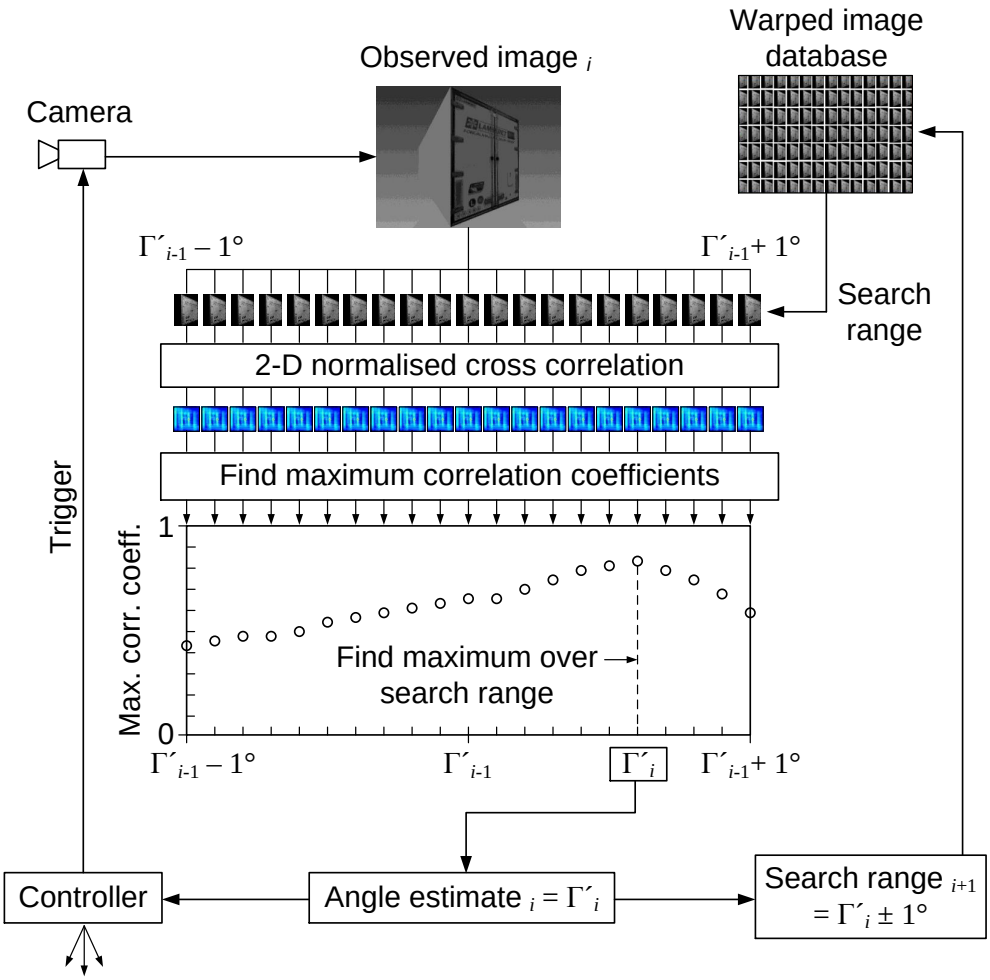


Fig. 3.10 Template-matching methodology. Search range = 2° (±1°), increment = 0.1°

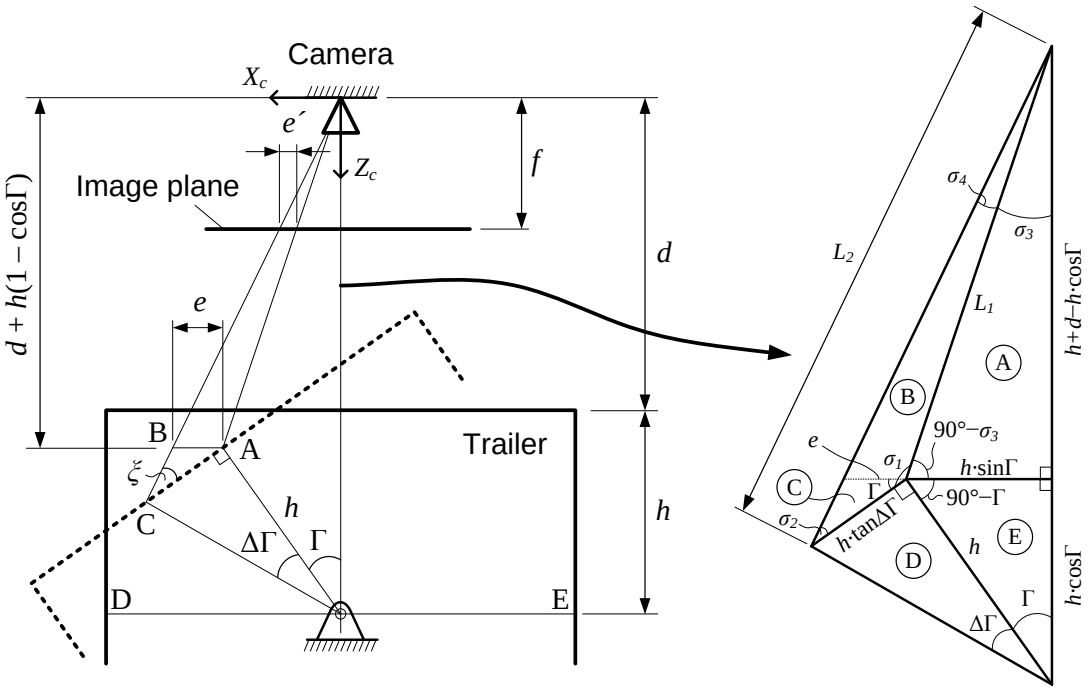


Fig. 3.11 Incremental lateral motion behaviour of the trailer in the image plane

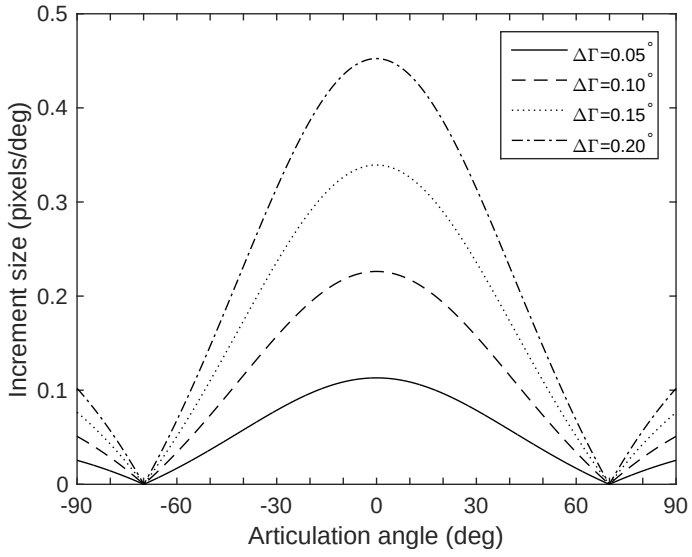


Fig. 3.12 Incremental motion behaviour Δu as a function of articulation angle Γ and articulation angle increment size $\Delta\Gamma$ (VGA resolution, d = 2.3 m, h = 1.2 m)

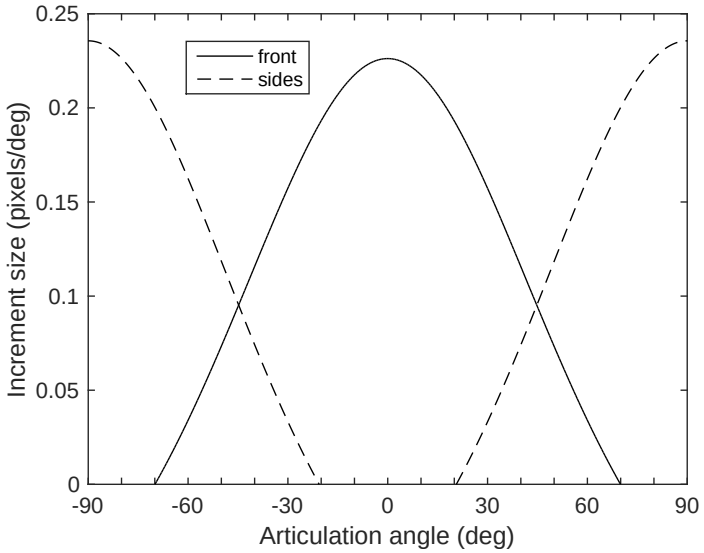


Fig. 3.13 Incremental motion behaviour Δu including the trailer sides (shown for a 0.1° articulation angle increment $\Delta\Gamma$).

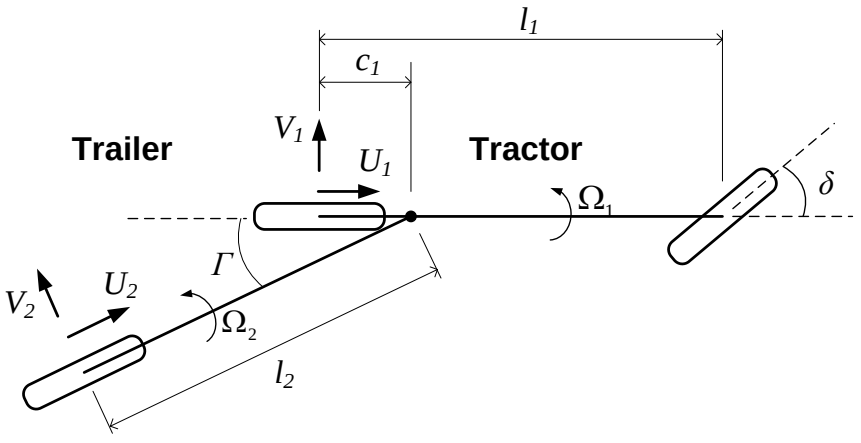


Fig. 3.14 Kinematic bicycle model of a tractor semi-trailer

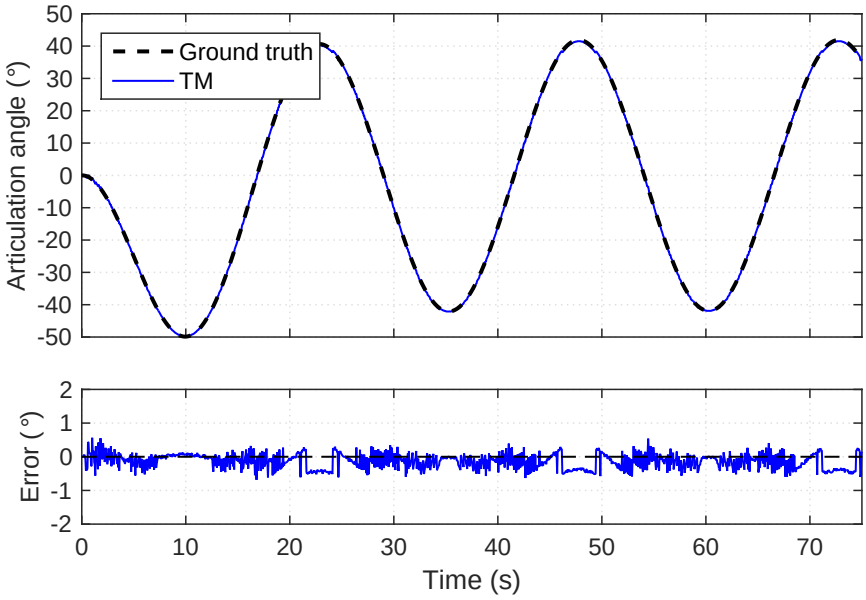


Fig. 3.15 Reference simulation results, $\Gamma \leq 50°$: no down-sampling, 0.1° search increment (TM only)

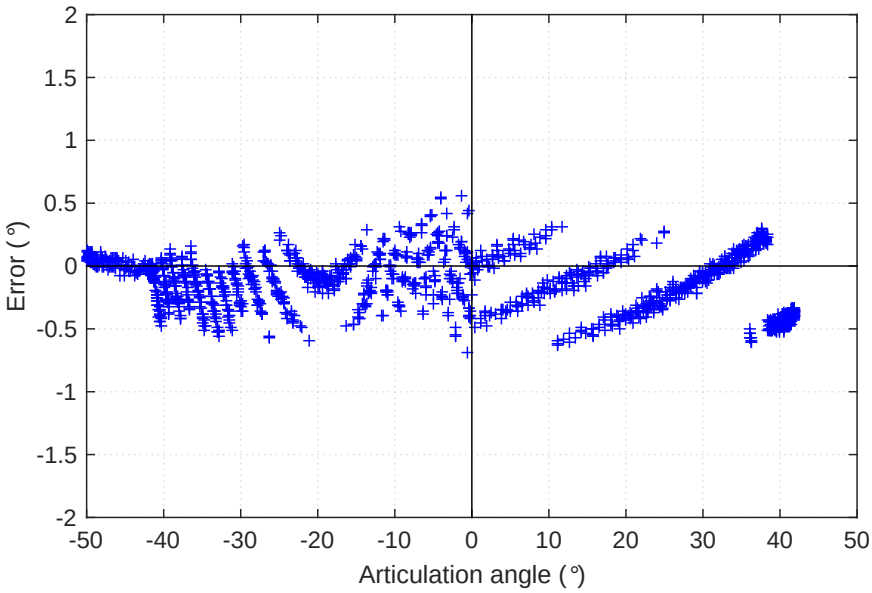
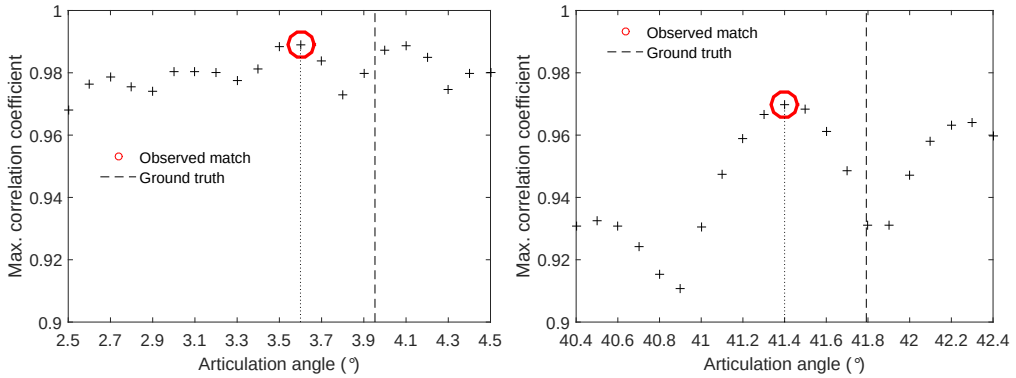
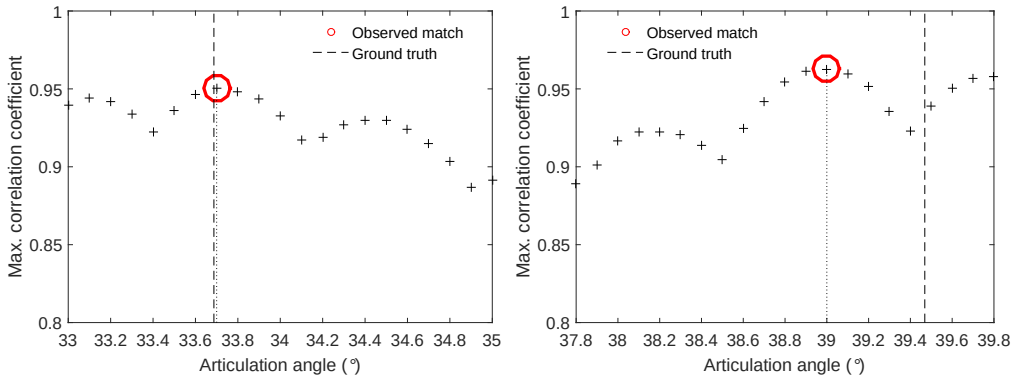


Fig. 3.16 Reference simulation results, $\Gamma \leq 50°$: no down-sampling, 0.1° search increment (TM only)



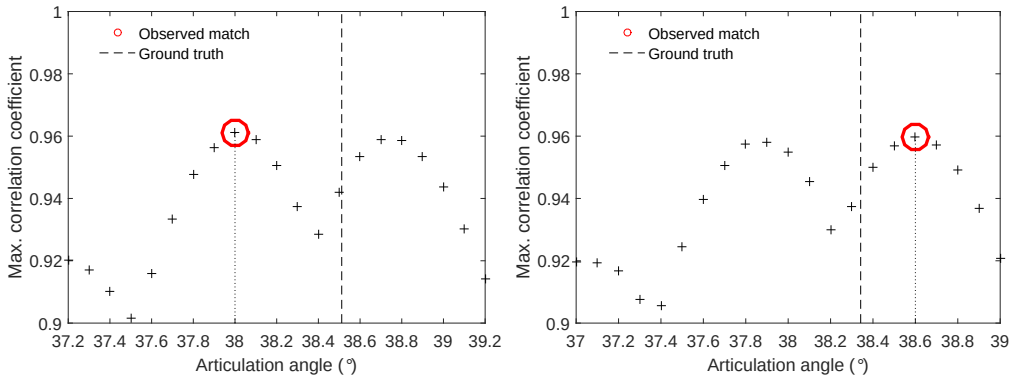
(a) (b)

Fig. 3.17 Variation of aliasing between low angles (a) and high angles (b)



(a) (b)

Fig. 3.18 A good match (a) versus a bad match (b): the effect of aliasing



(a) Before jump (t_i) (b) After jump (t_{i+1})

Fig. 3.19 A jump of estimated articulation angle between two peaks in sequential data points (TM).

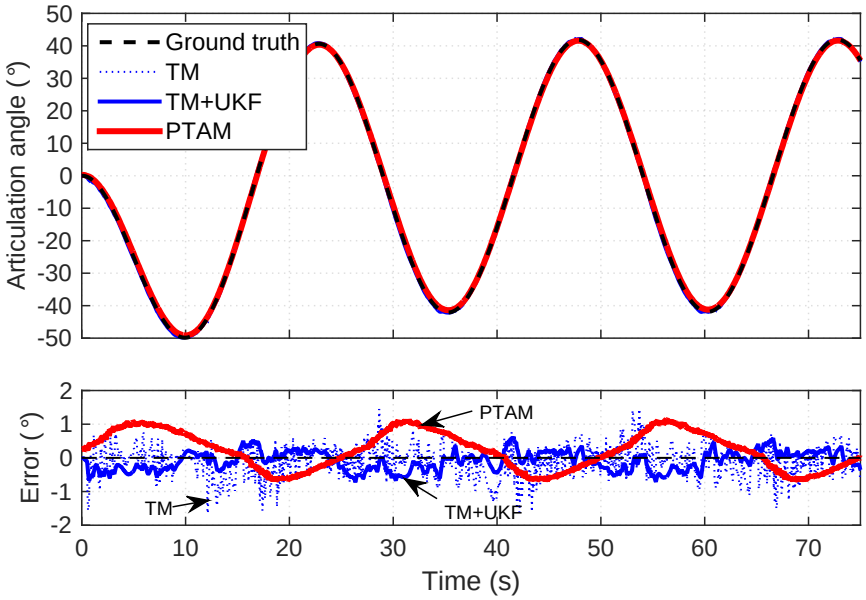


Fig. 3.20 Simulation results, $\Gamma \leq 50°$. Template-matching includes $1\times$ down-sampling and $0.2°$ search increment

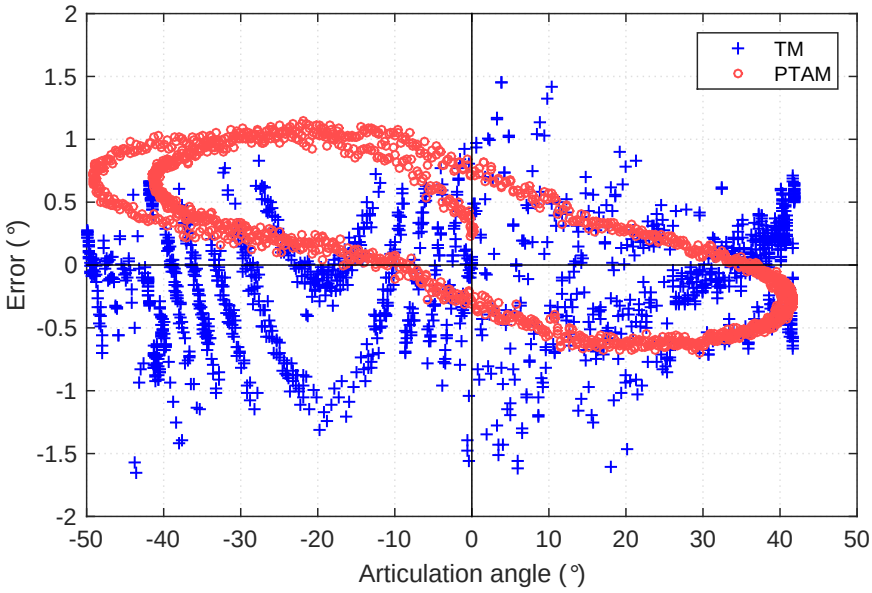


Fig. 3.21 Simulation results, $\Gamma \leq 50°$. Template-matching includes $1\times$ down-sampling and $0.2°$ search increment

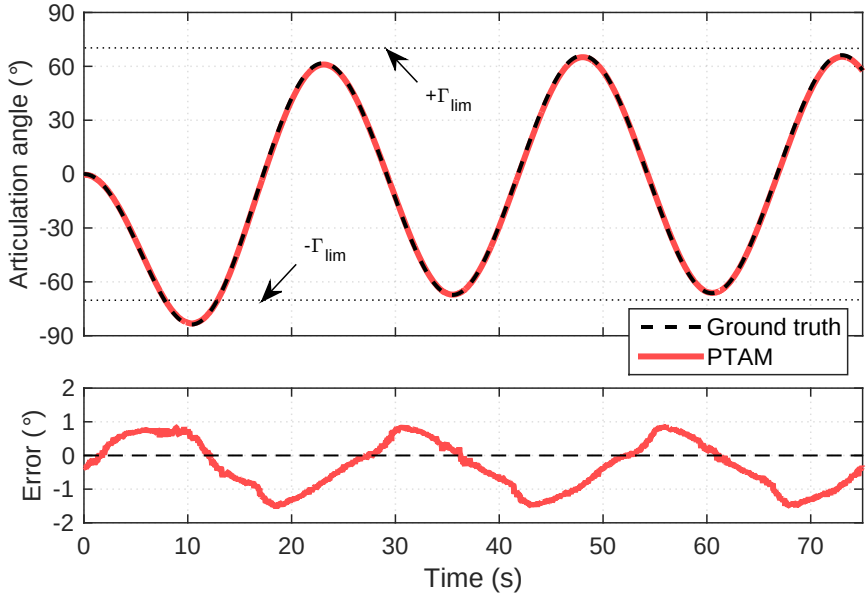


Fig. 3.22 Simulation results, $\Gamma \leq 90°$ (PTAM only)

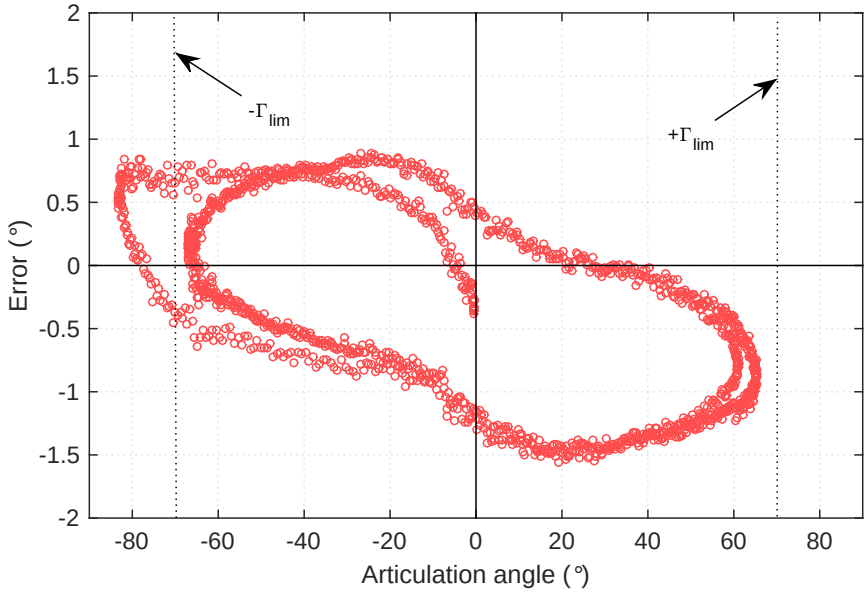
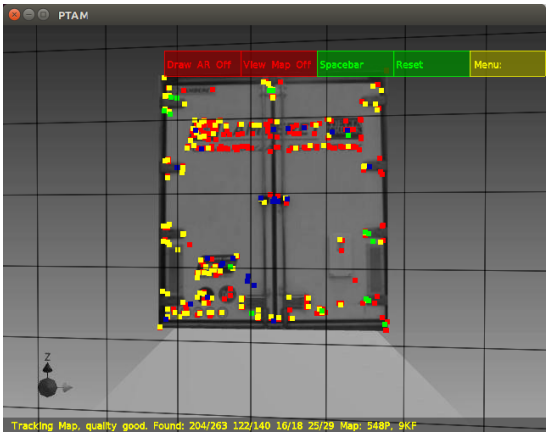
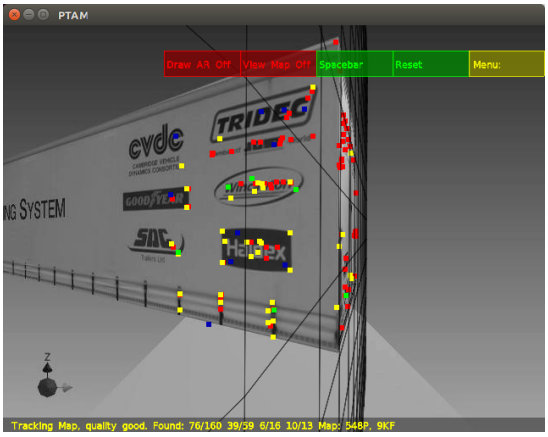


Fig. 3.23 Simulation results, $\Gamma \leq 90°$ (PTAM only)

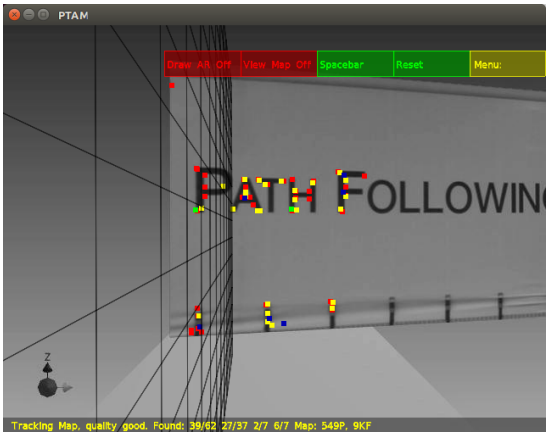
(a) Stereo initialisation



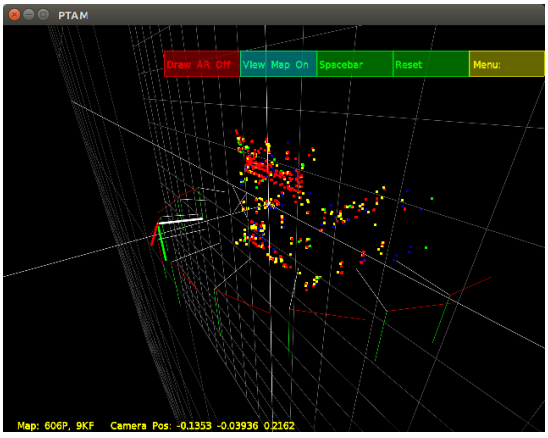
(b) Features detected on front



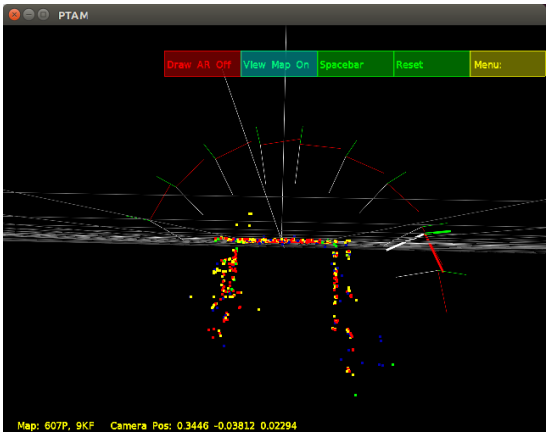
(c) Features detected on left side



(d) Features detected on right side

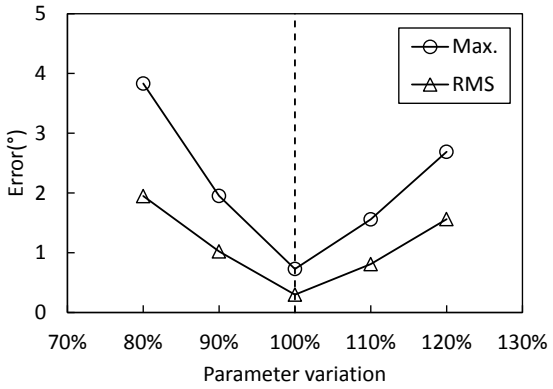


(e) 3-D feature map and keyframe locations

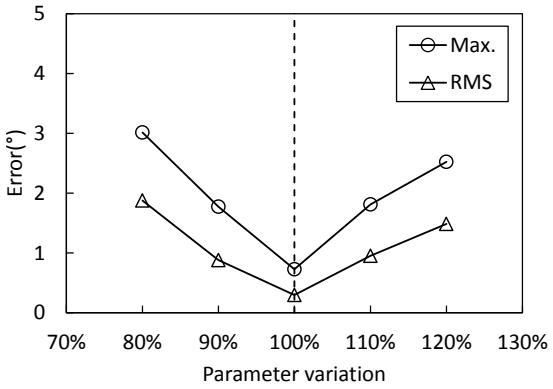


(f) Top view of 3-D feature map

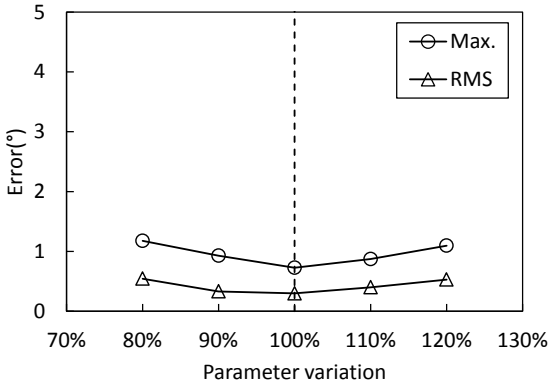
Fig. 3.24 PTAM feature tracking and initialisation



(a) Trailer front overhang (h)



(b) Hitch location (d, c_1)



(c) Trailer wheelbase (l_2)

Fig. 3.25 TM+UKF error sensitivity to $\pm 10\%$ and 20% variations in parameters ($\Gamma \leq 50^{\circ}$)

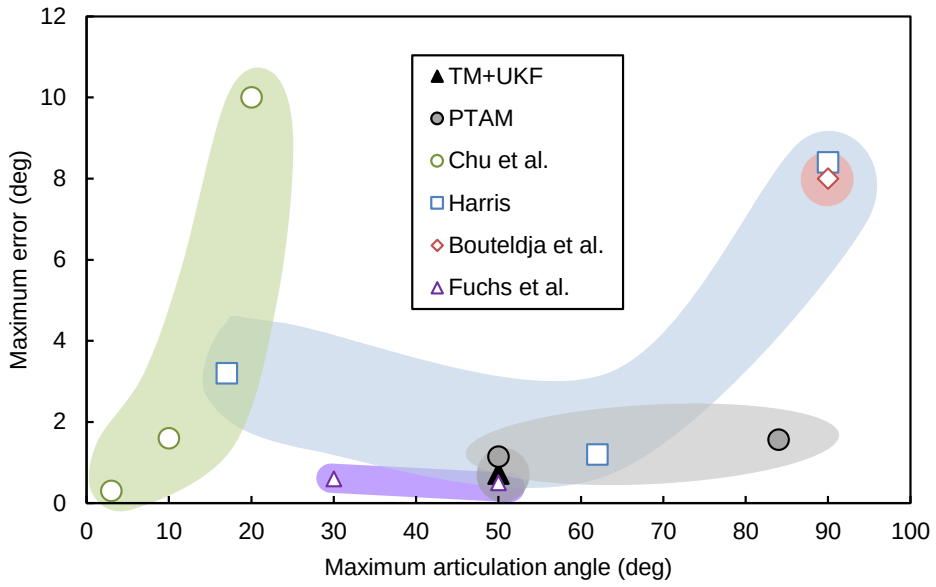


Fig. 3.26 Performance benchmark against published simulation data

Chapter 4

Field tests of the articulation angle sensor on a tractor semi-trailer

In the previous chapter, the template-matching and PTAM algorithms were developed and shown to exhibit good performance when simulated in an idealised CAD environment. In that environment there was no trailer motion outside of the yaw plane, lighting conditions were consistent and favourable, and cameras were perfect. In practice, trailers will exhibit pitch and yaw motion as the HGV travels on uneven road surfaces, lighting and weather conditions will be variable, and cameras will be imperfect. These will affect the performance of both these algorithms and this must be quantified to better understand the performance of the systems in reality.

To investigate, a full-scale tractor semi-trailer combination was instrumented and on-road vehicle tests were performed. This chapter details the experimental setup, testing procedure and results. It also includes an investigation into the error behaviour of the template-matching algorithm.

4.1 Experimental setup

4.1.1 Vehicle and instrumentation

Vehicle testing was carried out on an articulated vehicle combination consisting of a 4×2 Volvo tractor and a tri-axle, box-type 'B-link' trailer. The front axle of the tri-axle group was lifted, making this effectively a tandem-axle trailer. Although steerable, the trailer axles were not steered during testing. The vehicle is shown in Figure 4.1a with some important dimensions. Parameters d and h (Figure 3.3) were measured to be 886 mm and 1575 mm respectively.

In this experiment the B-link trailer is intended to represent a semi-trailer, and for the purposes of this testing there are no significant differences between the two, apart from the wheelbase which is slightly shorter than a conventional semi-trailer.

A Point Grey Flea3 USB 3.0 camera[1] was fitted to a bracket behind the tractor cabin, facing the front of the semi-trailer (Figure 4.1b). The camera was mounted centrally relative to the sides of the tractor cab and at an arbitrary height above the kingpin while maintaining a reasonable view of the front of the trailer. By eye, the optical axis of the camera was aligned with the lateral centre of the trailer. The lens used was a Fujinon YV2.8 $\times$ 2.8SA-2 wide-angle lens with an adjustable focal length of 2.8–8 mm. The focal length was set near its smallest value, giving a wide field of view.

The superstructure of the semi-trailer was a shipping container unit, resulting in a corrugated front face (*i.e.* not strictly planar). As the template-matching algorithm is limited to purely planar surfaces it was necessary to fix a planar surface to this face. To account for the loss in visual detail, artificial visual texture (of arbitrary design and position) was added on top of this. For the sake of the PTAM algorithm, some visual texture was also added to the sides of the trailer for higher articulation angles (see Figure 4.1c).

A schematic of the vehicle instrumentation is provided in Figure 4.2. Tractor speed and steer angle measurements were required for the UKF, and were acquired via a speed sensor on the tractor drive axle and a string potentiometer steer angle sensor on the tractor steer axle. A 'ground truth' articulation angle measurement was obtained with a VSE articulation angle sensor [66][2]. This was mounted on the kingpin of the semi-trailer (Figure 4.1d), and was calibrated before testing and was zeroed before each test.

Analogue sensor signals were logged using 'ICON' control computers (by Mechatronika Systems Ltd) on both tractor and trailer. Only their analogue-to-digital functionality was used here. The ICON has an output resolution of 4095 counts over a $\pm$10.4 V range, and the gain from measured articulation angle to ICON output was found to be 8.52°/V, giving a resultant reference articulation angle resolution of approximately 0.05°. This is sufficiently less than the target maximum error of 0.6°, and so deemed suitable.

Greyscale images were captured via USB 3.0 at 20 fps at a resolution of 640 $\times$ 480 with a dedicated Linux-based computer unit in the tractor cab. Image trigger signals and sensor signals were synchronised via a host PC running SIMULINK xPC Target. A CANbus link was used to communicate between the tractor and trailer ICONs and the xPC 'target' computer, as well as between the Linux computer and xPC target computer. A personal

[1] Model FL3-U3-13S2M-CS

[2] Recall that the VSE sensor is a trailer-based sensor, requiring trailer modifications at the kingpin and a non-standard communication link between tractor and trailer. This makes it unsuitable for commercial application, but its high precision makes it attractive for ground truth measurements for research and development purposes.

laptop computer was used as the xPC 'host' computer for data logging and general test control.

Sensors for steer angle, articulation angle and speed, along with the tractor and trailer ICON units and the xPC target computer, were already installed for previous work on these vehicle units. The camera, Linux computer and host computer were commissioned specifically for this work. Existing calibration data for the steer angle sensors were used, while new calibrations were carried out for the speed and articulation angle sensors. These sensors were calibrated before testing.

As discussed previously, the desired frame rate for this system is 10 fps. PTAM can achieve up to 20 fps, but the template-matching algorithm in its current form is not able to do so. Images were captured at 20 fps to best investigate PTAM's performance, but processed off-line at whatever frame rate was achievable by each algorithm. A efficient implementation of the template-matching algorithm (in embedded C-code for example) would likely have a comparable frame rate to PTAM. Therefore both algorithms were compared using identical image sequences.

4.1.2 Test scenarios

Testing was carried out on rough tarmac at Bourn airfield near Cambridge. Two trailer front faces were considered: a planar front (a predominantly flat surface) and a non-planar front (with a 3-D protrusion added). The planar case is shown in Figure 4.3a and Figure 4.3b. This case was applicable to both TM+UKF and PTAM algorithms. The non-planar front is shown in Figures 4.3c, 4.3d and 4.3e. Here a mock refrigeration unit (a common non-planar addition to trailer fronts) was affixed to the front of the trailer. The dimensions of the non-planar protrusion are shown in Figure 4.3f. This case was only applicable to PTAM.

For the planar case, two types of manoeuvre were carried out: a periodic step steer input with articulation angles up to 30° and a general driving manoeuvre (a pseudo-random set of turns). In both cases articulation angles were sufficiently small to maintain an acceptable view of the front of the trailer. Three tests of each scenario were conducted. An example of one of the periodic step steer inputs is shown in Figure 4.4.

For the non-planar case, three types of manoeuvre were carried out: two periodic step steer input cases with articulation angles of up to 30° and 50° respectively, and a general driving manoeuvre. Two tests of each case were conducted. The higher articulation angles in these tests enabled the effectiveness of PTAM to be assessed when the front trailer face disappears from view, when keyframes from the side of the trailer are obtained. An example image showing the visible side of the trailer is given in Figure 4.3e.

Vehicle speed was approximately constant at 5 km/h for the step steer tests and variable in the range 0–10 km/h for the general driving tests[3]. Steering input was manual in both cases, though controlled to an approximately repeatable steer amplitude and frequency in the step steer tests. The road surface was not necessarily level and was significantly pot-holed, resulting in some pitch and roll motion of the trailer during tests. Lighting conditions were mostly favourable, though light intensity was variable between tests and in some cases harsh shadows were evident on the trailer face.

Hereafter, the 30° and 50° periodic steer tests are denoted 'per30' and 'per50' respectively, and general driving scenarios are denoted 'gen'. The non-planar cases are indicated with the suffix "_3d".

4.2 Template-matching implementation

The template-matching algorithm was implemented in MATLAB as before with minor updates, including updated image cropping parameters, the addition of real-time distortion correction and updated noise covariances for the UKF. All other parameters were unchanged, including an image resolution of 640×480, a database search step size of 0.2°, a region of interest window of 60×40 pixels for locating template matches in sequential frames, and one level of down-sampling for NCC. Image sequences were processed offline on the same 3.2 GHz desktop computer used in the previous chapter (see Appendix A).

Observed images were automatically cropped to a size of 640×200. For the particular visual scenario used, this was found to have little effect on the performance of the algorithm, while significantly reducing computation time. Template images, generated from a cropped 640×200 datum image, were automatically cropped further as a function of articulation angle so as to exclude pixels with no information as a result of the warping process.

The camera was calibrated using the 'Camera Calibration Toolbox for MATLAB' [163], which uses the method of Zhang [161] and distortion model of Heikkilä [159] (Section 2.3). A checkerboard pattern with a 7×10 grid and a square size of 25.5 mm was used, and reprojection errors of less than 0.5 pixels were obtained. One of the calibration images is shown in Figure 4.5a with detected corners shown in Figure 4.5b. Camera calibration results are given in Appendix C. Distortion correction was performed in real-time at each frame using the OpenCV library [164]. An example image before and after the distortion correction

[3]The speed and articulation range considered in these tests are typical of an autonomous reversing manoeuvre for example. In comparison, jackknife control or combined braking and steering control would operate under significantly higher speeds. Higher speeds may introduce practical challenges such as vibration-induced motion blur, but would also typically present a smaller required measurement range of Γ.

step is shown in Figure 4.5c and Figure 4.5d respectively. The computational cost of this step was small compared to other steps in the template-matching process.

For the Unscented Kalman Filter, the measurement noise covariance, W, was determined experimentally to be $(1.45^\circ)^2$. The process noise, Q, was $(0.06^\circ)^2$ as before, tuned through trial and error on a few initial datasets.

A datum image for the TM+UKF algorithm was collected only once for the planar trailer and was used for all six tests. This was obtained during a preliminary straight line driving manoeuvre. Care was taken to avoid a test with harsh shadows for the datum image. The tests therefore embodied the realistic scenario of using a datum image which was obtained under potentially different lighting conditions.

4.3 PTAM implementation

As described in Section 2.3, PTAM uses a simplified distortion model with only one (radial) distortion coefficient. The open-source implementation of the PTAM algorithm includes a camera calibration module which incorporates this model, and this was used to recalibrate the camera for PTAM processing using the same 7×10 checkerboard. Camera calibration results are provided in Appendix C.

Full image resolution was used (no cropping). No additional changes to the standard PTAM algorithm from [123] were made. Results for PTAM were processed offline on the 2.4 GHz laptop computer used previously (see Appendix A for specifications).

4.4 Results: planar trailer front

Results for each of the six vehicle tests with the planar trailer front are given in Figure 4.6. Unfiltered template-matching results ('TM'), template-matching results with the UKF ('TM+UKF'), and PTAM results are shown. Although the overall template-matching algorithm includes the UKF, the unfiltered results are included as these provide useful insights to accompany the following discussion. Error signals (relative to the VSE sensor) are also shown. A summary of the maximum and RMS errors is given in Table 4.1 and a comparison of error performance between TM+UKF and PTAM is shown in Figure 4.7.

Average frame rates in the region of 6–10 fps were achieved for TM+UKF, slightly less than the 10 fps target but deemed acceptable for a prototype MATLAB implementation of the algorithm. Variations in frame rate were due to the manner in which the template images were cropped according to articulation angle, resulting in smaller images at larger articulation angles. Frame rates for PTAM were 20 fps, well within the requirement.

Some general observations may be made from Figure 4.6. Firstly it is clear that the unfiltered TM results are noisy as was observed in the CAD simulations. A portion of this noise may be attributed to aliasing, the main source of noise in the idealised CAD simulations. However, unlike in the CAD simulations, another contribution probably comes from small wavelength and low amplitude undulations in the road surface during testing.

Where larger amplitude and higher wavelength variations in the road surface were encountered these are clearly observable in the image sequences as roll and/or pitch motion of the trailer. Examples of this can be seen in Figure 4.6c at around t = 48–49 s and Figure 4.6d at around t = 85–92 s.

In the first case a sharp outward roll motion is evident from the image sequence, resulting in a seemingly increased articulation angle (as the trailer face becomes narrower). This agrees with the observed TM signal which over-estimates Γ in this period.

In the second case, an increase in trailer pitch is evident as the vehicle slows to a stop, with Γ close to zero. The result is that the trailer face appears closer to the camera, yielding a mismatch between the observed images and the saved datum image (and template images near 0°) which will be at a smaller scale. Visually, this is neither an apparent increase or decrease in Γ, and should arbitrarily result in incorrect template matches in one direction or the other. In this case a negative error is observed.

Consistent regions of error are clear in Figures 4.6a, 4.6b and 4.6c between the error peaks at A and B and from C to D (shown on the plot of each error signal). These errors are also clear in Figure 4.6d at 53–60 s and Figure 4.6f at 10–20 s, but are less obvious because the variations in articulation angle are not consistent and repeatable; although the errors appear to exist at articulation angles consistent with the periodic manoeuvre results. These errors could not be correlated with observed variations in roll and pitch motion in the image sequences. Rather, these will be shown to be a particular subset of the errors as discussed in the following section and are denoted the 'recurring' component of the error trend.

In all cases the UKF sufficiently corrects for higher frequency noise due to aliasing and road roughness. In some cases it also reduces the effect of lower frequency variations such as in Figure 4.6c at t = 48–49 s, provided the time interval is sufficiently small. On average it has the effect of reducing maximum errors from 5.88° to 3.92°, and RMS errors from 1.84° to 1.32° (see Table 4.1 and Figure 4.7).

Superior results were achieved when datum images specific to each test run were used, but the results presented here offer a more realistic indication of real-world performance with only one datum image obtained for the given trailer.

Figure 4.8 shows details of the variation of the maximum correlation coefficient during several 'snapshots' of the 'per30 C' test. Each plot shows the results for 11 image comparisons

per time step (a search window of $\pm 2^\circ$ with a search increment of 0.2° gives a total of 11 image comparisons). The maximum correlation coefficient in the search range is highlighted for each, which represents the selected 'match' and the estimated value of Γ.

For each of these matches (*i.e.* the one circled data point out of 11 in Figure 4.8), the associated correlation map and overlaid image match is shown in Figures 4.9 and 4.10 respectively. In Figure 4.10, the bounding box shows the border of the template image, and an averaged image is shown in the region of overlap. A number of observations may be made from each of these snapshots (in order of appearance):

1. Figures 4.8a, 4.9a, 4.10a: This is a good match at a small Γ, with a smooth variation in maximum correlation coefficients (Figure 4.8a), a well defined peak in the correlation map (Figure 4.9a) and an easy image match (Figure 4.10a). Good performance is expected near the datum in the absence of disturbances.

2. Figures 4.8b, 4.9b, 4.10b: This result comes from one of the distinct peaks in errors just after the transition from $-\Gamma$ to $+\Gamma$. The match result looks good (Figure 4.10b) and the variation in correlation coefficients is smooth. However, the maximum correlation coefficient is at the extreme of the search range (Figure 4.8b) indicating that the actual maximum is likely outside of the search range. After a few more time steps the search window will quickly move towards higher Γ until the peak is again within the search range. Although the error is high, it seems the template match has performed well, indicating some other source of the error.

3. Figures 4.8c, 4.9c, 4.10c: This is in the region of steady positive errors where the template-matching results appear smooth and consistent but a moderate error is exhibited. This is therefore possibly an indication of a physical phenomenon.

4. Figures 4.8d, 4.9d, 4.10d: This data point was taken from one of the points associated with a transient trailer pitch motion, which is not associated with the other recurring error trends. It is clear from 4.10d that the best template match has not quite captured the observed image correctly. Significant offset between template and observed images is clear, notably in the upper half of the template image. This error correlates well with the transient pitch motion of the trailer. The trailer has pitched forward, rotating the trailer face outward at the top. The warped template images do not have this pitch component inherent in them.

A final observation to make is that a second peak in the correlation map of Figure 4.9d is evident in the lower right. This is where the background scene has become visible

(see Figure 4.10d) and illustrates the need for the region of interest window discussed in Section 3.2.2.

Overall, from Figure 4.7, PTAM yields the superior performance with the smallest average maximum and RMS errors of 1.83° and 0.73° respectively. The PTAM signal is smooth and displays only a repeatable increase in error as a function of articulation angle, especially for $-\Gamma$. This is likely attributable to the simpler distortion model of PTAM, the inaccuracies of which would become more pronounced near the borders of the image, giving correlations with images at higher Γ. It is also clear that PTAM is robust to the pitch and roll variations discussed above.

A qualitative view of the PTAM algorithm for the "per30 C" test is shown in Figure 4.11. Figures 4.11a and 4.11b show the initialisation step and features detected. As expected, most detected features are located on the clear visual features of the attached visual texture, though some have been detected on the bare trailer face.

All features are tracked effectively as Γ increases as shown in Figure 4.11c and also under variations in lighting intensity as shown in Figure 4.11d (most features are retained from Figure 4.11b). The generated scene map of feature points is shown in Figure 4.11e where the reference plane has been fixed to the features on the planar trailer face (the plane is also shown in Figures 4.11a to 4.11d), and the path of relative camera motion is shown as an arc in front of this plane. A view perpendicular to the plane of the trailer front is shown in Figure 4.11f in which some erroneous out-of-plane feature points can be observed. The algorithm was shown to be robust to these outliers.

4.4.1 Error trends

As mentioned previously, certain errors in the unfiltered template-matching results (the 'recurring' errors) could not be attributed to any obvious pitch or roll motions of the trailer. These errors are now investigated in detail.

The recurring error behaviour is best illustrated by examining variations of the error with articulation angle, as shown in Figure 4.12 for each of the six tests. PTAM errors are included for comparison. The repeatability of the error trends between periodic and general driving tests is confirmed by overlaying results from all six tests as shown in Figure 4.13.

For the purpose of discussion the TM error trend (Figure 4.13a) can be separated into two components. Ignoring for a moment the range $-15^\circ \leq \Gamma \leq 10^\circ$, there appears to be a small, approximately linear variation in errors with Γ. This is the first component and is denoted the 'smooth' trend component. At small values of Γ there is a second component which appears to override the first. In the range $-5^\circ \leq \Gamma \leq 5^\circ$ there is a strong linear trend peaking at the error maxima of approximately -3° and 5°. This is denoted the 'zero-crossing' error trend

component, given that the peak errors appear either side of the zero-crossing region. In the remaining ranges $-15^\circ \leq \Gamma \leq -5^\circ$ and $5^\circ \leq \Gamma \leq 30^\circ$ there is a transition region between the two components.

In comparison, the PTAM error trend (Figure 4.13b) exhibits a smooth variation with articulation angle, noted previously to be likely due to PTAM's simplified camera distortion model. The trend is similar to the smooth component of the template-matching trend in that it is an approximately positive variation with Γ. Although the magnitude of this gradient is not completely comparable to that of template-matching, the common trend may be indicative of a common physical phenomenon, in addition to the camera distortion model.

It should be noted that a steady state offset of magnitude $+2.7^\circ$ was observed in the raw output of PTAM, which was accounted for in all tests as part of the calibration process. It is interesting that this offset is comparable to the offset of the smooth component of the template-matching trend. The source of this offset is unclear: it may be indicative of a camera yaw offset, 'crabbing' motion of the trailer[4], an irregularity in the structure of the trailer superstructure, or characteristic of the initialisation process. This steady state error could be zeroed out easily.

The zero-crossing component of the template-matching error trend (at $-5^\circ \leq \Gamma \leq 5^\circ$) is not at all replicated in the PTAM trend, suggesting that this is purely a symptom of the template-matching algorithm itself, possibly in response to a physical phenomenon.

To investigate the source of the template-matching errors in more detail, it is useful to revisit the underlying assumptions of the system model. These were as follows:

1. Trailer motion is constrained to rotation in the yaw plane about a vertical axis located at the fifth wheel. This implies that there is no out-of-plane trailer rotation (pitch and roll) and no translation of the kingpin relative to the fifth wheel (*i.e.* no fifth wheel lash).

2. The z-axis of the camera is perpendicular to the vertical axis of the fifth wheel, and is parallel to the yaw-plane. This means that there is no constant yaw or roll offset between the camera and trailer.

3. The geometry of the system in terms of d and h is known accurately.

4. The camera properties are known accurately.

[4]Crabbing motion is when axle misalignment causes a discrepancy between the vehicle's heading and its direction of motion. For rigid vehicles this would require a small steer input at the front axle to maintain straight forward motion. For a trailer with misaligned axles, it could result in a small steady articulation angle when the lead vehicle is following a straight path.

5. The datum image is obtained with the trailer at exactly zero articulation angle.

Variations from these ideal conditions may exist in multiple forms. The most likely scenarios include:

1. misalignment in the mounting of the camera, yielding a constant rotational and/or translational offset;
2. a constant pitch or roll offset between the tractor cab and semi-trailer as a result of axle ride height setup, cab suspension setup or kingpin lash;
3. transient variations in pitch or roll between the tractor cab and semi-trailer due to undulations in the road surface;
4. inaccuracies in the measurements of d and h; and
5. an inaccurate datum image, with inherent offset in pose.

Variations in pitch and roll (either constant or transient) would have a dual effect, firstly through deviations in the assumed yaw-plane motion model, and secondly though variations in the values of d and h from their assumed nominal values. Furthermore, a constant offset in pitch also raises the possibility that the nominal values of d and h were measured with some inherent pitch offset, and that the assumption that $\mathbf{N} = [0, 0, 1]^\top$ for the datum image is no longer valid.

Given the transient pitch and roll motion observed during testing due to road undulations, it is also possible that some rotation and/or translation is inherent in the datum image. This would also affect the assumed values of d, h and $\mathbf{N}$.

4.4.2 Pitch, roll and translation

A useful feature of the PTAM algorithm is that it outputs a complete 3-dimensional estimate of the relative pose. It can therefore be used to quantitatively investigate pitch, roll and translational motion between the camera and trailer, and hence help to evaluate some of the above hypotheses about the template-matching algorithm.

PTAM's accuracy in articulation measurement has been shown to be reasonable, with average maximum and RMS errors of 1.82° and 0.73° respectively. Its accuracy in pitch and roll estimation should be comparable, however it should be noted that no reference signals are available to confirm this or to zero these signals.

Figure 4.14 shows the relative pitch and roll motion outputs of PTAM, for all six tests overlaid. Both signals have been processed so as to represent pitch and roll motion in a

reference frame fixed to the trailer. This is covered in Section 3.4.3. A pitch offset of approximately 2–2.5° is evident at small Γ, with some variation with Γ. Transient variations in pitch in individual tests are also visible, showing variations of up to approximately ±1°.

Roll motion is approximately zero for small Γ; an offset here would be meaningless as the roll-orientation of the plane fixed to the trailer front during initialisation is arbitrary. A negative gradient trend with Γ is evident, as are transient variations of up to approximately ±1°. A pitch or roll offset with no variation with Γ would indicate a pure pitch or roll in the trailer reference frame. Given that there are some variations with Γ, this is either indicative of an offset in the camera mounting, or of some other physical phenomenon.

It should be recalled that an offset of approximately 2.7° was observed in the raw data for Γ, which should be incorporated into the discussion at this point. A yaw offset may also be indicative of a lateral camera offset. A lateral offset in the mounting of the camera, and subsequent adjustment of the camera's yaw orientation so as to centre the trailer in the field of view, would result in both a yaw offset and lateral offset in the camera mounting. Similarly, a roll offset in the trailer might result in a camera yaw offset, as it's yaw orientation is adjusted to account for the lateral shift of the trailer due to its roll angle. Therefore the yaw offset, if a result of camera mounting error, may be indicative of either trailer yaw offset or trailer roll offset.

4.4.3 CAD variations

To investigate the effects of out-of-plane rotations and translations in a controlled environment, the original CAD model was updated to match the dimensions of the test vehicle (including measured values of d and h). To replicate the visual characteristics of the real tests, an undistorted image of the planar front of the real trailer was superimposed onto the front of the CAD trailer. The side of the trailer was similarly reproduced. Pitch and roll degrees of freedom were added to the trailer and the camera location was made to be adjustable. The updated CAD model is shown in Figure 4.15.

Variations in trailer roll, pitch, yaw and camera yaw and offset were applied to the CAD model, and simulations were rerun with a basic sinusoidal articulation variation over a comparable range of Γ. The resultant error trends as a function of Γ were then compared to those of the vehicle tests. A comparable error trend from the variation of one variable would suggest that that variable may have been the underlying cause of errors in the vehicle tests.

Results for a subset of these tests are given in Figure 4.16. These are overlaid on the TM results for all six vehicle tests for comparison. A baseline CAD simulation with no offsets or irregularities is shown in Figure 4.16a, and the trend is largely as expected though with some minor deviations near $\Gamma = 0°$. The noise component of the CAD errors is significantly

smaller than the vehicle tests. The noise component for the CAD simulations is most likely from aliasing which will also be inherent in the vehicle data; the additional noise in the vehicle data is assumed to be due to transient pitch and roll variations.

The addition of 2° pitch angle to the trailer (Figure 4.16b) produced an error trend similar to the vehicle tests in both the smooth component and the zero-crossing component. However, an offset between CAD and vehicle tests of approximately 1° is evident in the smooth component of the trend. The addition of trailer roll (Figure 4.16c) produced a convincing trend for positive Γ only, but with magnitudes more comparable with vehicle tests than the case of trailer pitch. It is possible that the trailer roll angle could have changed direction in the transition at $\Gamma = 0^\circ$, in which case the CAD data for $+\Gamma$ in Figure 4.16c would be reproduced symmetrically about the vertical axis but inverted for $-\Gamma$, which would produce a trend similar to the vehicle test data shown in the background for all Γ.

Camera yaw offset (Figure 4.16d) did not replicate vehicle test trends particularly well. Camera lateral offset (Figure 4.16e) produced a somewhat convincing trend for $-20^\circ \leq \Gamma \leq 0^\circ$, but not at all for $+\Gamma$. The combined effect of trailer roll and camera yaw (Figure 4.16f) did not in general replicate the error trends from the vehicle tests.

Given these observed trends, along with the observed PTAM pitch offset of 2–2.5° and the equally notable 2.7° PTAM yaw offset, in combination with physical considerations and observed image sequences, it was concluded that trailer pitch and camera offset (accompanying trailer yaw) were the most likely causes of the error trends. These are investigated in more detail in the following section.

4.4.4 Correction models

With the goal of removing the contribution of these effects from vehicle test results, correction models were developed to account for both trailer pitch and camera offset. These were assessed using the CAD simulations.

Consider the general effects of irregular motion on the assumed motion model of the template-matching algorithm, particularly that of the image warping step. From Equations 3.16 and 3.17, the affected parameters could be one or more of the following: d, h, $\mathbf{N}$, R, $\mathbf{T}$, or the datum image could be inaccurate. (The camera calibration matrix can be neglected in this discussion).

It should be noted that whether an offset is inherent in the camera mounting or in the trailer, this will change the parameters affected. For example, pitch offset in the camera mounting would affect only the value of $\mathbf{N}$, the assumed orientation of the plane relative to the camera optical axis in the datum image. However, a positive pitch in the trailer about the

kingpin would also affect d as the trailer face pitches closer towards the camera, and possibly also h, depending on whether the plane of articulation changes or not.

The proposed trailer pitch correction model is illustrated in Figure 4.17, showing a simplified side view of the trailer-camera system. Assuming articulation to still occur in a plane aligned with the trailer yaw plane (*i.e.* h is unchanged), only parameters d and $\mathbf{N}$ would be affected. d is the distance from the camera to the datum plane *normal to the datum plane* (not to the optical axis).

For a pitch angle of θ, d would be corrected to d_{new} as indicated in the diagram. The component of the change in d in the z-direction (along the optical axis), Δd_z, and the resultant corrected value of d, d_{new}, may be derived by geometry and shown to be:

$$\begin{aligned} \Delta d_z &= [H\sin\theta - h(1-\cos\theta)] + \tan\theta\,[h\sin\theta + H(1-\cos\theta)] \\ d_{new} &= (d - \Delta d_z)\cos\theta \end{aligned} \tag{4.1}$$

where H is the height of the optical axis above the pitch centre, assumed to be at the kingpin. For the test vehicle H was measured to be 1750 mm.

Accompanying this would be a change in $\mathbf{N}$ to reflect the non-zero pitch angle of the datum plane. This can be incorporated by pre-multiplying the nominal value of $\mathbf{N}$ by the pitch component of a rotation matrix as follows:

$$\mathbf{N}_{new} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\sin\theta \\ \cos\theta \end{bmatrix} \tag{4.2}$$

Consider now the effect of camera offset. In Equation 3.15, the x, y and z components of $\mathbf{T}$ were derived under the assumption of perfect alignment between the camera and hitch point. If we now assume an offset of Δx, additional terms will be introduced to the x- and z-components of $\mathbf{T}$. The proposed offset correction model is illustrated in Figure 4.18, showing a simplified top-down view of the trailer-camera system. The corrected value of $\mathbf{T}$ may be deduced from geometry and shown to be:

$$\mathbf{T}_{new} = \begin{bmatrix} (h+d)\sin\Gamma \\ 0 \\ (h+d)(1-\cos\Gamma) \end{bmatrix} + \begin{bmatrix} \Delta x(\cos\Gamma - 1) \\ 0 \\ \Delta x\sin\Gamma \end{bmatrix} \tag{4.3}$$

The effect of a camera yaw offset to realign the optical axis with the centre of the trailer could be accounted for by suitably adjusting $\mathbf{N}$ in a manner similar to the pitch correction above. This was not considered here.

The effectiveness of the pitch correction was assessed by implementing it in the CAD simulation with 2° pitch. Uncorrected and corrected simulation results are shown in Figures 4.19a and 4.19b respectively. It is clear that the smooth error component has been corrected, but that the zero-crossing trend at small Γ has been largely unaffected.

As part of an investigation into the cause of this, the edge detection template-matching method (see Section 3.2.2) was implemented in place of the standard template-matching method. Results for the 2° pitch case with no pitch correction but with edge detection are shown in Figure 4.19c. Interestingly, this has resulted in an almost complete correction of the zero-crossing component of the error trend. Results with both pitch correction and edge detection are shown in Figure 4.19d, showing an almost perfect reproduction of the baseline CAD case shown in Figure 4.16a.

The offset correction model was applied to the 50 mm camera offset CAD simulation. Uncorrected and corrected simulation results are shown in Figures 4.20a and 4.20b respectively. Again, it is clear that only the smooth component has been corrected, and not the zero-crossing component. Figure 4.20c shows the result with both offset correction and edge detection—again a close reproduction of the original unperturbed CAD scenario, although some outliers are evident at small Γ.

The correction models have been validated and shown to adequately correct for pitch and camera offset in the smooth component of the error versus Γ trend. However, it was also shown that the zero-crossing component of the trend appears to be a unique characteristic of the NCC template-matching method.

In this case the addition of edge detection to the template-matching step appeared to correct for this behaviour. To explore this further, edge detection was used in reprocessing of the vehicle test data. Although errors seemed to be reduced in general, a small disturbance such as a small pitch or roll motion would result in a sudden and irreversible drift in error rendering the algorithm significantly less robust. This behaviour reflects the findings of Section 3.2.2. As a result, edge detection was not included for further processing and analysis.

Pitch correction was found to have the dominant effect on vehicle test results, and no conclusive improvements were exhibited by the inclusion of camera offset correction (of any reasonable magnitude). It was found that a pitch correction of 2.2° gave the best improvement of the results, removing the high Γ component of the errors. As expected, the large errors at small Γ remained.

4.4.5 Corrected results

Errors as a function of Γ are shown in Figure 4.21 with and without pitch correction. The remaining errors at small Γ are still clear, though reductions in the steady errors at higher Γ

are evident. The correction effect evident in Figure 4.21 is small, though it is clear that the 'smooth' trend component has been flattened to some degree.

A summary of maximum and RMS errors for the individual tests is given in Figure 4.22, with tabulated results in Table 4.2. The pitch correction reduced the average maximum error and average RMS error by 0.47° and 0.26° respectively in the TM+UKF results. This yielded an average maximum error of 3.45°, about twice that of PTAM, and an average RMS error of 1.06°, about 50% more than that of PTAM.

The 'gen A' vehicle test exhibited the poorest TM and TM+UKF performance, largely due to its prolonged duration in the 5–10° range of Γ between 50 and 60 s. In all cases except 'per30 B' and 'per30 C', PTAM errors are consistently smaller than those of TM+UKF. In the exceptions it is likely that the initialisation of PTAM could have been improved.

The largest effect on performance has been shown to be a symptom of the normalised cross correlation template-matching method used. The underlying mechanism of this effect has not been clearly determined, however if this could be addressed then the performance of the template-matching system would be notably improved.

4.4.6 Parameter sensitivity

The assumed trailer dimensions to which the base algorithm is sensitive are the front overhang, h, and the hitch location, $h+d$ (Figure 3.4). Although the hitch location is technically a tractor parameter, it is moveable to some extent and so is included for consideration here. In addition, the UKF introduces sensitivities to the assumed trailer wheelbase and the number and spread of the trailer axles, and is also sensitive to the hitch location.

For the parameter sensitivity study, the 'per30 C' test was processed using the TM+UKF algorithm with pitch correction. Parameter variations of $\pm$10% and $\pm$20% of trailer front overhang, hitch location, trailer wheelbase, and axle spread were investigated. Variations of axle spread were deemed to be an adequate representation of variations in number of axles as well.

Sensitivity results are shown in Figure 4.23 in the form of maximum and RMS errors. A parameter value of 100% represents the baseline value as used in the previous results. Figures 4.23a and 4.23b demonstrate some sensitivity to trailer front overhang and hitch location, with between 60% and 110% increases in RMS error due to 20% parameter variations. It should be noted however that these parameter variations are substantial, equivalent to a $\pm$2 m change in wheelbase, and $\pm$0.3 m change in hitch location. These are likely at the upper end of expected variations in practice. Sensitivity to $\pm$10% parameter variations are notably smaller at between 15% and 37%.

Maximum errors are less indicative of overall performance (as these can occur in response to irregular disturbances), and so the observed variations in maximum error are less informative. This is partially evidenced in the reduction in maximum errors with parameter variations in two instances. Nevertheless, variations in maximum error ranged from 11% and 52% for 20% parameter variations. Overall, RMS and maximum errors remained within 1.7° and 3.7° respectively for 25% variations in trailer overhang and hitch location.

Considering the UKF, Figures 4.23c and 4.23d demonstrate that sensitivity to trailer wheelbase and axle spread is comparatively minor. A 31% increase in RMS error was observed for a 20% reduction in trailer wheelbase. Only a 5% increase in RMS error was observed for an equivalent increase in wheelbase. For standard European semi-trailers (for example), vehicle weight and dimension regulations coupled with pressure to maximise legislated loading capacity result in fairly consistent designs in terms of length, wheelbase and overhang. Therefore, given the relatively minor sensitivity to the basic parameters, a default set of parameters could be assumed in practice with only relatively minor variations in performance. Sensitivity to axle spread is negligible in this case, with a maximum 1% increase in RMS error with a 20% variation in axle spread.

In practice, it is reasonable to assume that some basic trailer data may be available, or that trailer dimensions within a particular country or state would be fairly constrained to a small range of variations. However, small variations from the assumed values may occur. Tractor data such as wheelbase can be assumed known.

In comparison, PTAM should not be sensitive to any of these variations, as none of these parameters are required or assumed. These parameters would have a small effect on the range of trailer yaw rates during turning, and possibly the maximum articulation angles experienced, but these effects are expected to be negligible.

Furthermore, it was observed that the PTAM results exhibited well-filtered behaviour, with only a simple decaying velocity filter which makes no assumptions about trailer geometry. It is hence possible that a simpler filter could suffice for the TM algorithm, nullifying the requirement for trailer wheelbase and axle spacing information.

4.5 Results: non-planar trailer front

PTAM results for the non-planar trailer are shown in Figure 4.24. Qualitatively, performance is good with repeatable error fluctuations with Γ. Tracking is accurate and continuous up to larger values of Γ at which the front face of the trailer is obscured and features on the side of the trailer are detected and added to the scene map.

The initialisation step and tracked features are shown for the 'gen_3d B' test in Figures 4.25a and 4.25b. About a quarter of initial feature points are located on the non-planar refrigeration unit, and so the reference plane is still fixed to the largely planar front trailer surface. However, as there was more uncertainty in the automatic placement of this plane (relative to the planar tests), PTAM was zeroed at the beginning of each non-planar test.

At moderate Γ some feature points become occluded behind the refrigeration unit (Figure 4.25c) which has no effect on performance. At even higher Γ, the front of the trailer is almost entirely obscured. However, as the side of the trailer starts to become visible, features on the side of the trailer are detected and added to the scene map with no effect on the smoothness or accuracy of tracking (Figure 4.25d).

The scene map and relative camera motion is shown in Figures 4.25e and 4.25f (top view). Non-planar feature points are visible both on the refrigeration unit and on one side of the trailer.

Maximum and RMS errors are given in Table 4.3, and summarised in Figure 4.26. Average maximum and RMS errors were 2.73° and 1.19° respectively. Comparing results for the 30° tests with those of the planar tests shows comparable performance. However, the tests at larger Γ exhibited greater errors. It was shown previously that PTAM errors were approximately proportional to Γ, possibly the result of the camera distortion model. This seems to be the case here.

For more severely non-planar trailer geometries, some additional caution may be necessary during the initialisation phase to ensure a correct reference camera pose. The horizontal mounting of the camera for this work was not a strict requirement for the operation of either template-matching or PTAM algorithms, provided the initial pose of the camera was known. Variations in this would allow for more complex trailer scenarios to be included. For example, for flat-bed trailer types, a downwards-tilted camera with PTAM could be used, so as to work both with laden and unladen trailers. Similar mounting considerations could give rise to workable solutions for trailers with less regular fronts such as car-carriers. The hitching of a new trailer could be automatically detected, and a new initialisation process could detect the new trailer geometry, regardless of the trailer type.

4.6 Performance benchmark

A comparison of the final results with the published state-of-the-art is shown in Figure 4.27. Maximum errors observed in individual vehicle tests are shown as a function of maximum articulation angle tested. Showing this as a function of articulation angle accounts for some tests which may have only been performed over a small articulation range. It is clear that both

the TM+UKF and PTAM algorithms exhibit significantly improved accuracy over existing systems. TM+UKF is however limited to a maximum articulation angle defined by the trailer geometry. PTAM is observed to offer the best performance over a theoretically unlimited range, and without the restriction of planar trailer fronts.

4.7 Conclusions

1. The TM+UKF and PTAM algorithms were assessed in tests on a full-scale tractor semi-trailer combination. Both planar and non-planar trailer front face scenarios were investigated.

2. The TM+UKF algorithm achieved average RMS and maximum errors of 1.32° and 3.92° respectively at 6–10 fps for a planar trailer front at articulation angles up to 37°. Error trends highlighted a pitch offset between camera and trailer. Correcting for this resulted in RMS and maximum errors of 1.06° and 3.45° respectively.

3. The NCC image comparison method was shown to result in peak errors at low articulation angles which could not be corrected. This would need to be addressed in future work on the TM method.

4. The PTAM algorithm achieved average RMS and maximum errors of 0.73° and 1.82° respectively at 20 fps for the same planar front tests. Additional tests for a non-planar trailer front at articulation angles up to 55° yielded average RMS and maximum errors of 1.19° and 2.73° respectively, also at 20 fps.

5. Transient pitch and roll motion caused by road irregularities was shown to have a detrimental effect on the accuracy of the TM+UKF algorithm, but not on the PTAM algorithm.

6. The TM+UKF algorithm was shown to be moderately sensitive to assumed values of hitch location and trailer front overhang, but with RMS errors remaining within 1.7° for 20% variations in these parameters. Sensitivity to trailer wheelbase and axle spacing was shown to be small, suggesting that a simple filter which does not require vehicle information may be sufficient.

7. Both algorithms were shown to have superior accuracy to the current state-of-the-art in vehicle tests at the range of articulation angles tested. Errors were in excess of the target values, but could be reduced through further refinement in future work.

4.8 Tables and figures

Table 4.1 Results summary: planar vehicle tests

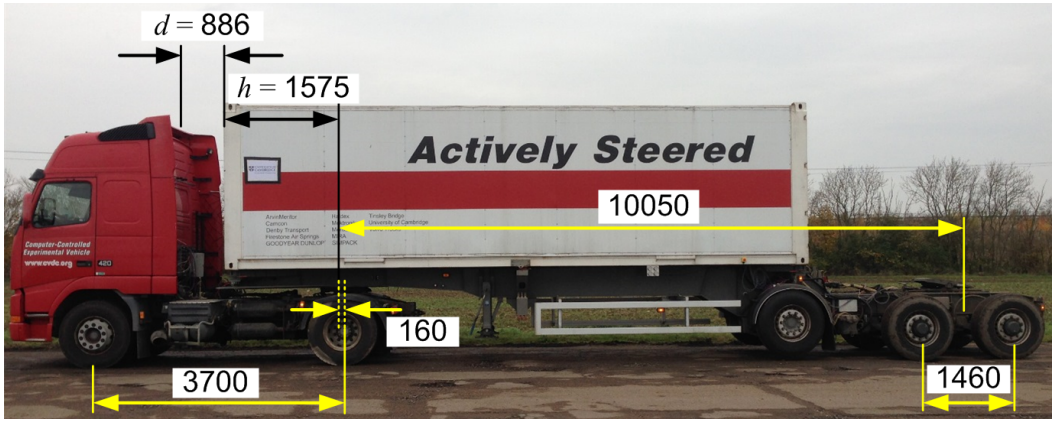
		TM		TM+UKF		PTAM	
	Γ_{max} (°)	ε_{max} (°)	ε_{RMS} (°)	ε_{max} (°)	ε_{RMS} (°)	ε_{max} (°)	ε_{RMS} (°)
per30 A	31	4.21	1.73	4.21	1.40	2.30	0.81
per30 B	29	5.64	1.61	2.37	1.04	2.57	1.04
per30 C	29	5.83	1.82	3.39	1.14	1.75	0.82
gen A	37	7.22	2.07	7.24	2.04	1.20	0.58
gen B	33	5.58	2.01	3.67	1.32	1.51	0.70
gen C	23	6.82	1.81	2.63	0.97	1.62	0.40
Ave. of tests		5.88	1.84	3.92	1.32	1.82	0.73

Table 4.2 Results summary: planar vehicle tests (with correction)

		TM		TM+UKF		PTAM	
	Γ_{max} (°)	ε_{max} (°)	ε_{RMS} (°)	ε_{max} (°)	ε_{RMS} (°)	ε_{max} (°)	ε_{RMS} (°)
per30 A	31	4.01	1.36	3.69	1.04	2.30	0.81
per30 B	29	5.79	1.25	2.42	0.86	2.57	1.04
per30 C	29	5.48	1.34	2.70	0.75	1.75	0.82
gen A	37	6.72	1.73	6.75	1.79	1.20	0.58
gen B	33	4.68	1.53	2.83	1.02	1.51	0.70
gen C	23	6.82	1.68	2.30	0.88	1.62	0.40
Ave. of tests		5.58	1.48	3.45	1.06	1.82	0.73

Table 4.3 Results summary: non-planar vehicle tests (PTAM only)

	Γ_{max} (°)	ε_{max} (°)	ε_{RMS} (°)
per30_3d A	31	2.26	1.14
per30_3d B	28	1.71	0.87
per50_3d A	50	3.46	1.58
per50_3d B	44	2.74	1.07
gen_3d A	35	2.44	1.20
gen_3d B	55	3.77	1.31
Ave. of tests		2.73	1.19



(a)

(b)

(c)

(d)

Fig. 4.1 Test vehicle and instrumentation: (a) test vehicle (dimensions in mm), (b) camera behind the cab, (c) visual texture on trailer front, (d) VSE articulation angle sensor viewed from the top of the kingpin assembly

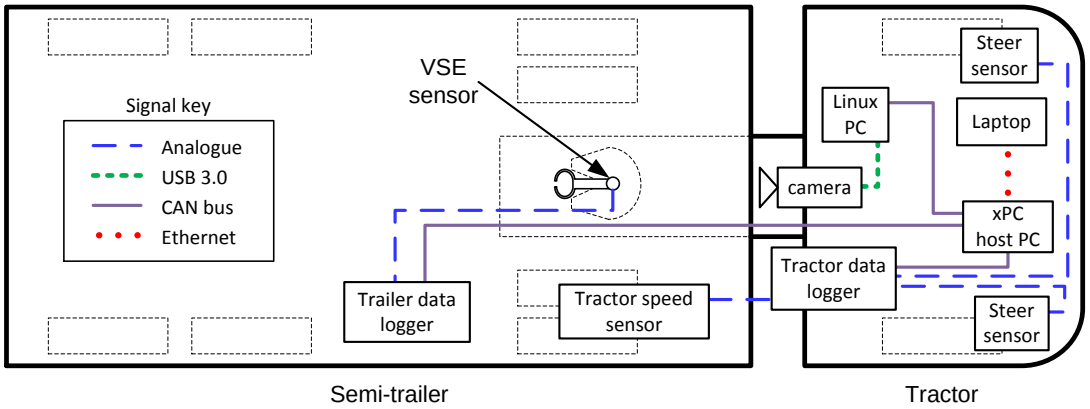
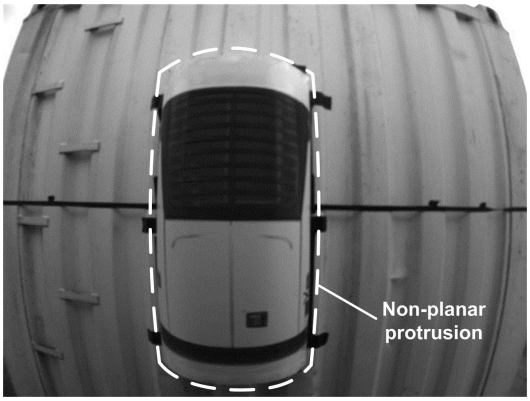


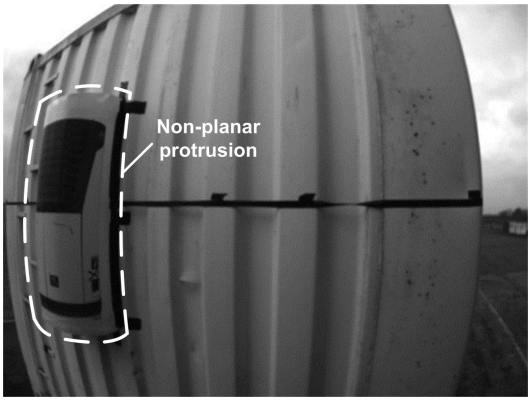
Fig. 4.2 Vehicle instrumentation layout

(a) Planar front near $\Gamma = 0°$

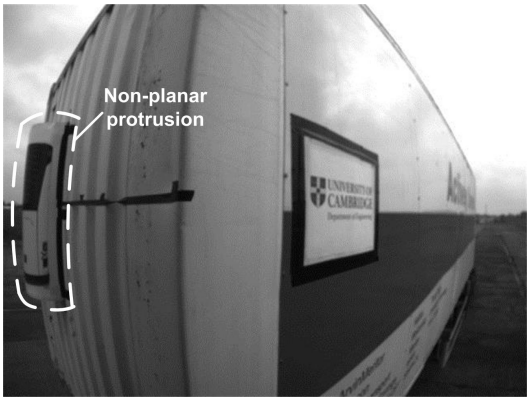
(b) Planar front at non-zero Γ



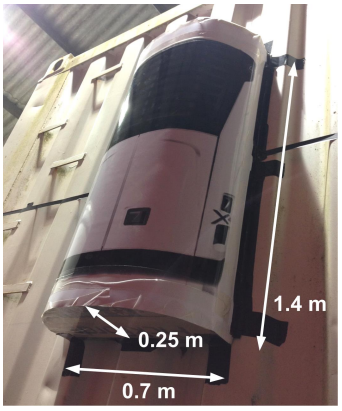
(c) Non-planar front near $\Gamma = 0°$



(d) Non-planar front at non-zero Γ



(e) Non-planar front at large Γ



(f) Artificial refrigeration unit

Fig. 4.3 Trailer front scenarios

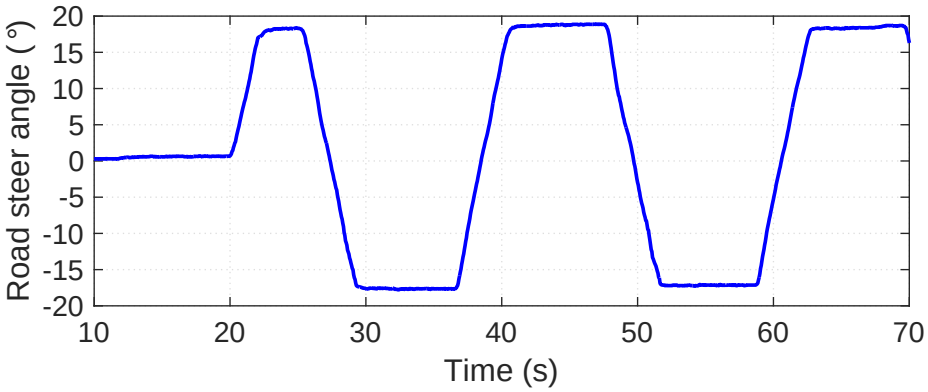
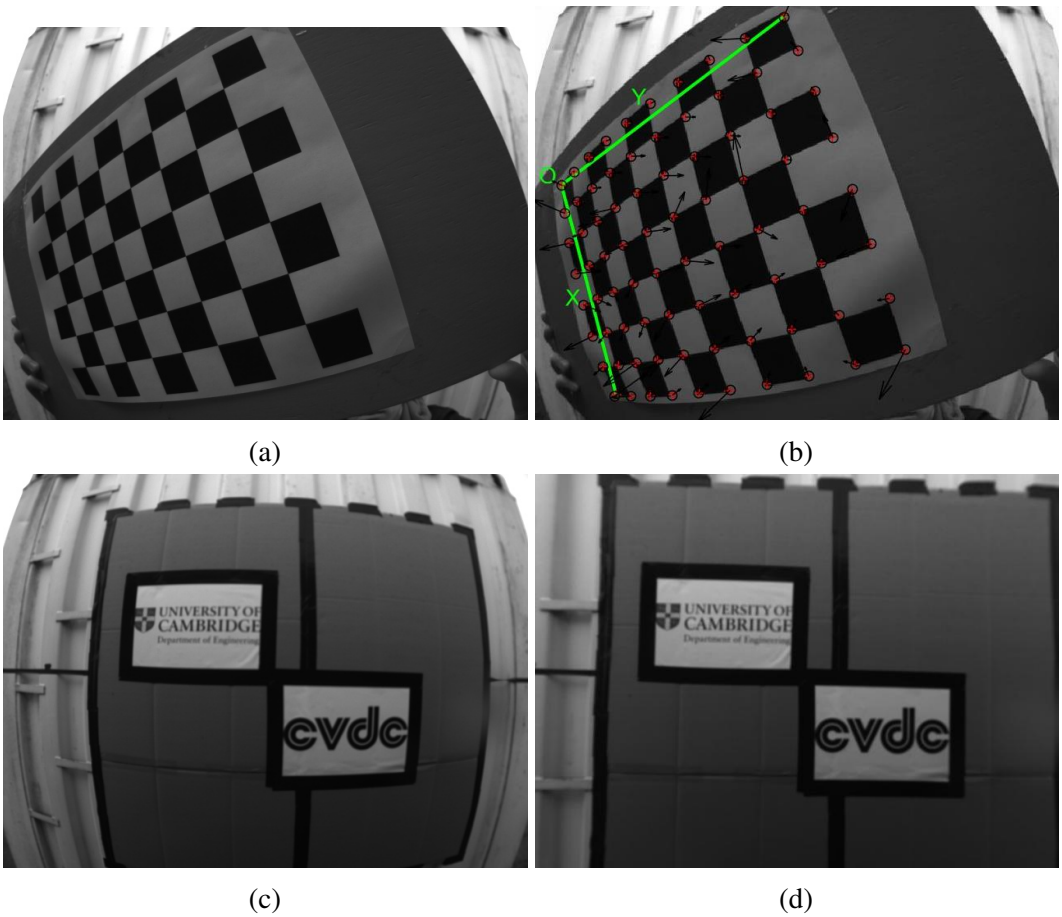


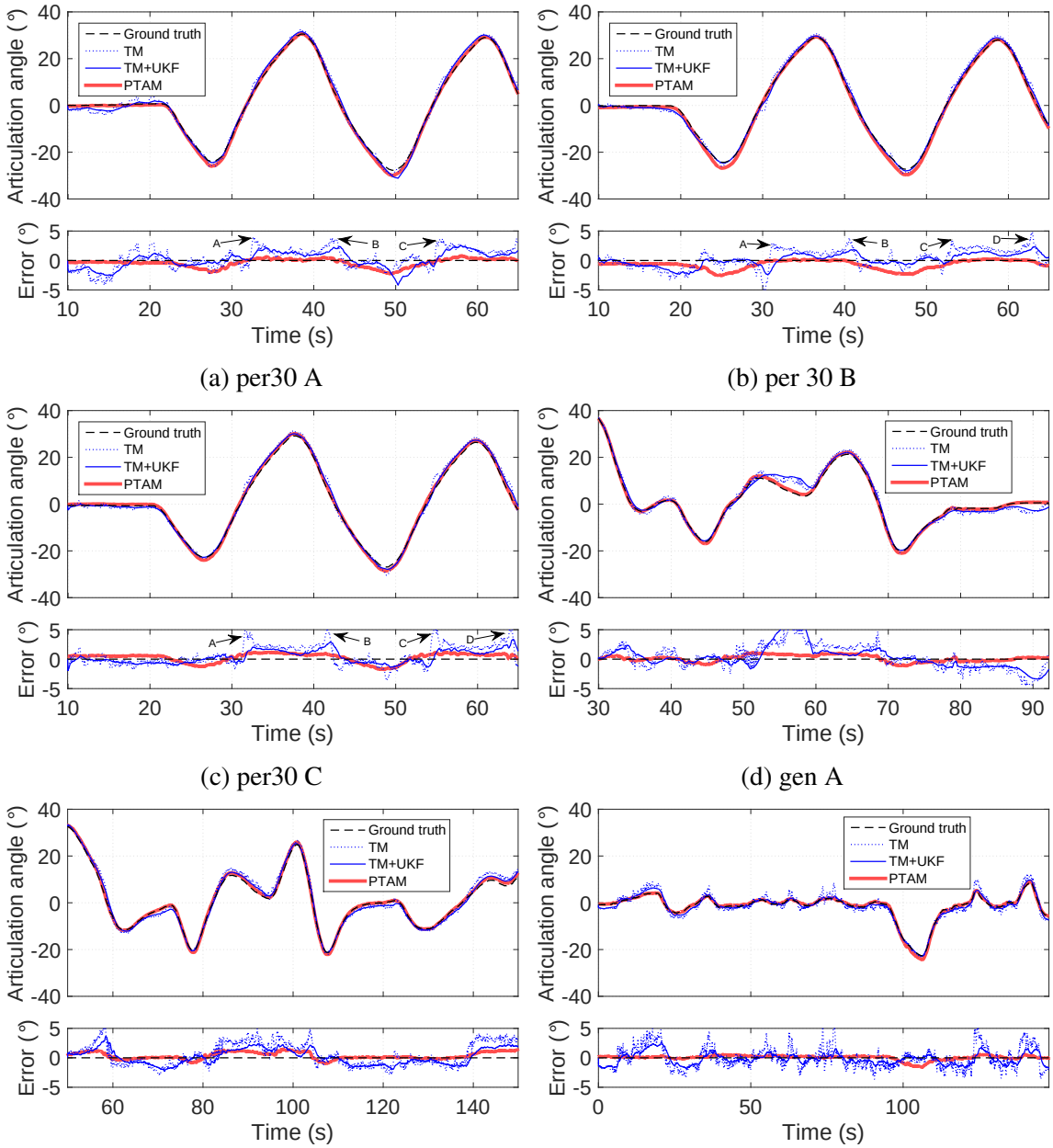
Fig. 4.4 Example periodic step steer input



(a) (b)

(c) (d)

Fig. 4.5 Camera calibration: (a) image of checkboard, (b) checkerboard corners detected, (c) disorted trailer image, (d) undistorted trailer image



(a) per30 A

(b) per 30 B

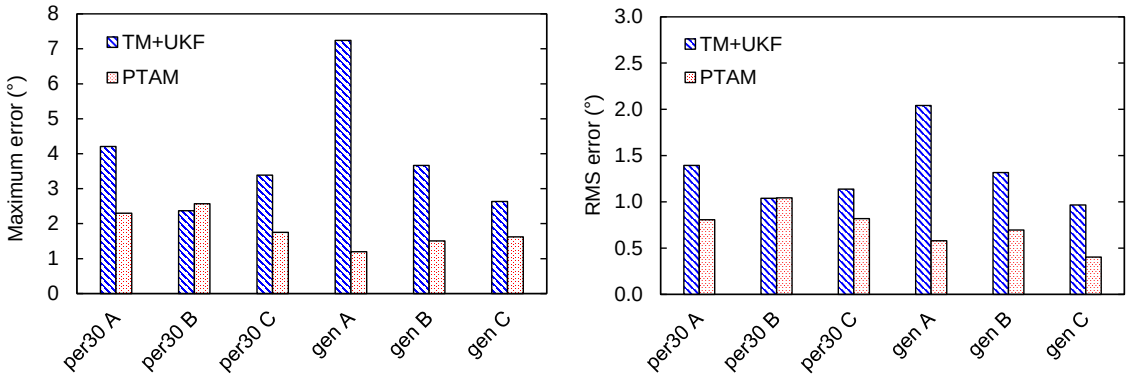
(c) per30 C

(d) gen A

(e) gen B

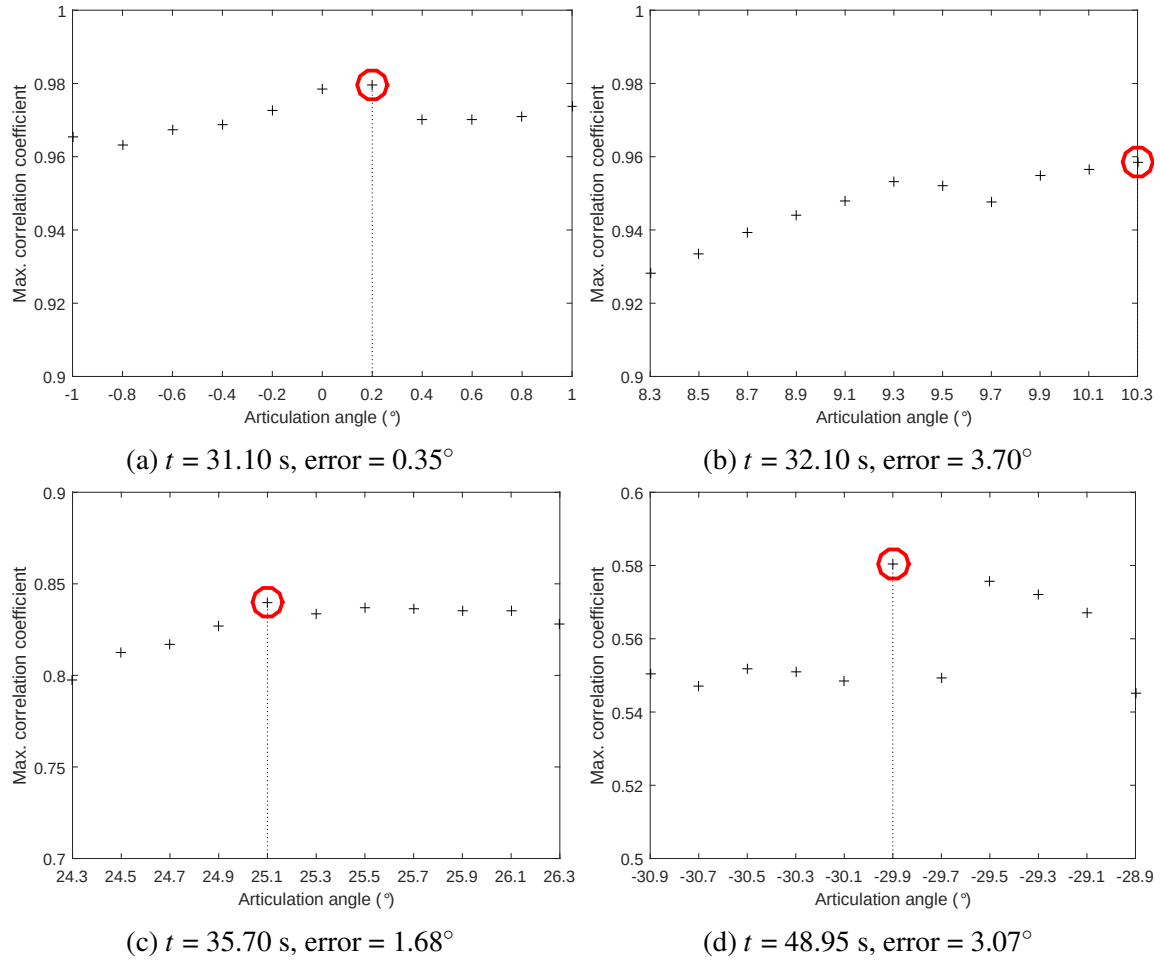
(f) gen C

Fig. 4.6 Vehicle test time histories, planar trailer



(a) Maximum errors (b) RMS errors

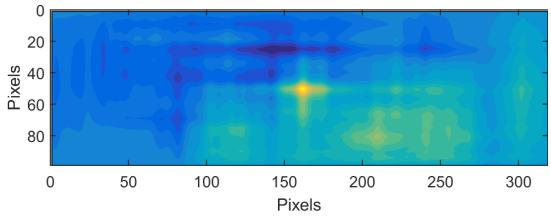
Fig. 4.7 Vehicle test errors, planar trailer front



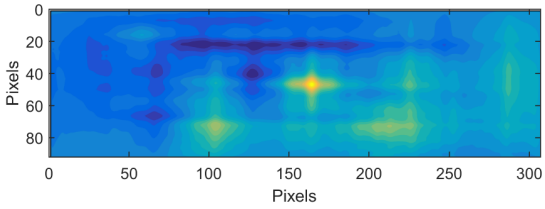
(a) t = 31.10 s, error = 0.35° (b) t = 32.10 s, error = 3.70°

(c) t = 35.70 s, error = 1.68° (d) t = 48.95 s, error = 3.07°

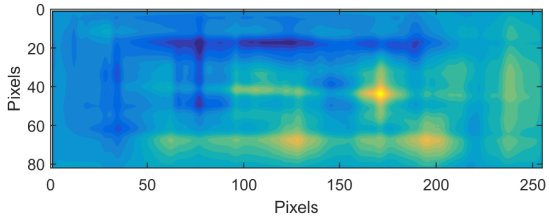
Fig. 4.8 Variations of maximum correlation coefficient over the search range for individual time steps, 'per30 C' vehicle test.



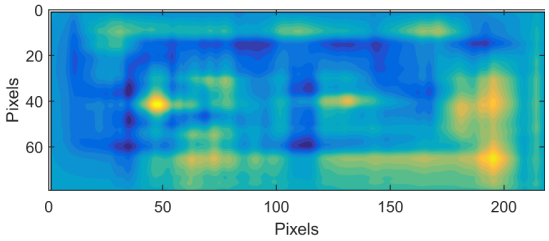
(a) t = 31.10 s, error = 0.35°



(b) t = 32.10 s, error = 3.70°



(c) t = 35.70 s, error = 1.68°

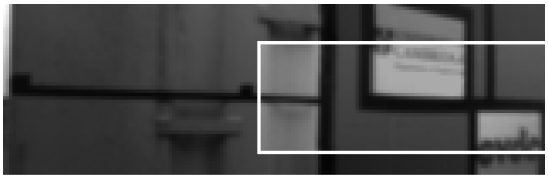


(d) t = 48.95 s, error = 3.07°

Fig. 4.9 Cross correlation maps for individual time steps, ‘per30 C’ vehicle test

(a) t = 31.10 s, error = 0.35°

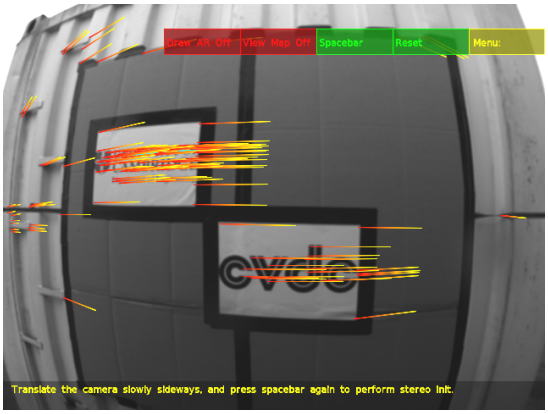
(b) t = 32.10 s, error = 3.70°



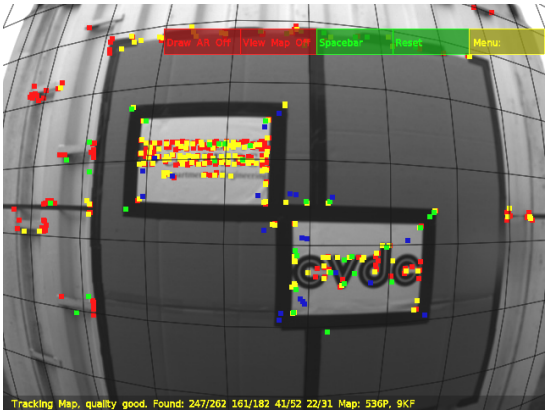
(c) t = 35.70 s, error = 1.68°

(d) t = 48.95 s, error = 3.07°

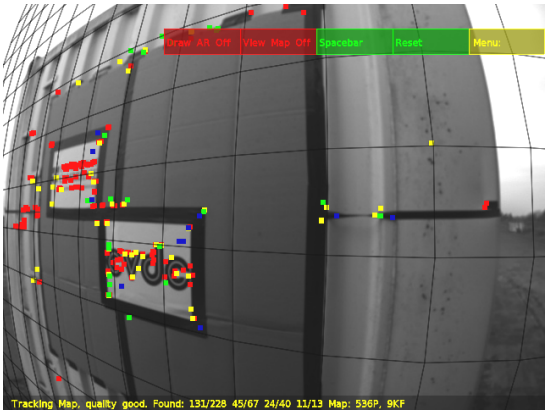
Fig. 4.10 Template match results for individual time steps, ‘per30 C’ vehicle test.



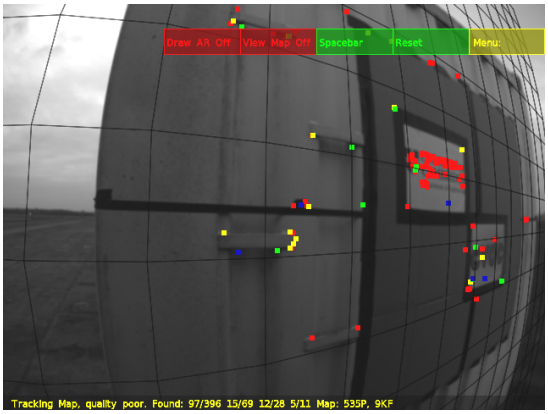
(a) Initialisation



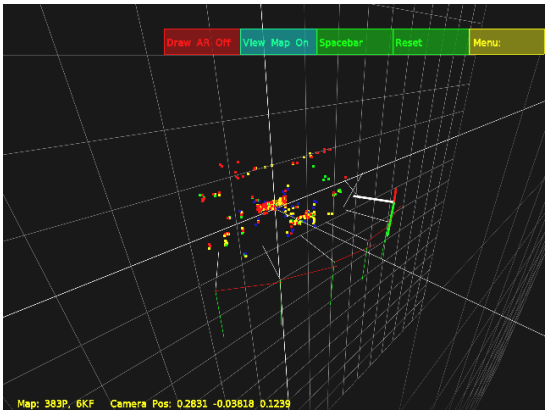
(b) Detected features



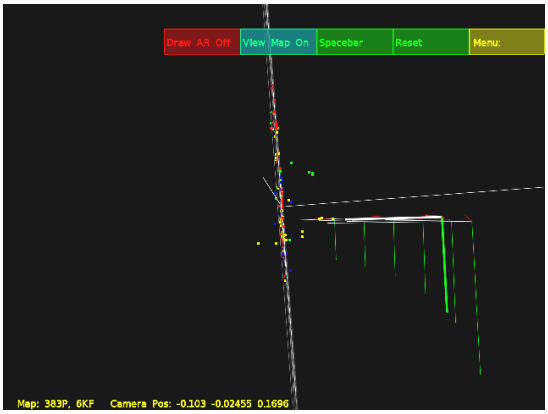
(c) Moderate Γ



(d) Reduced light intensity

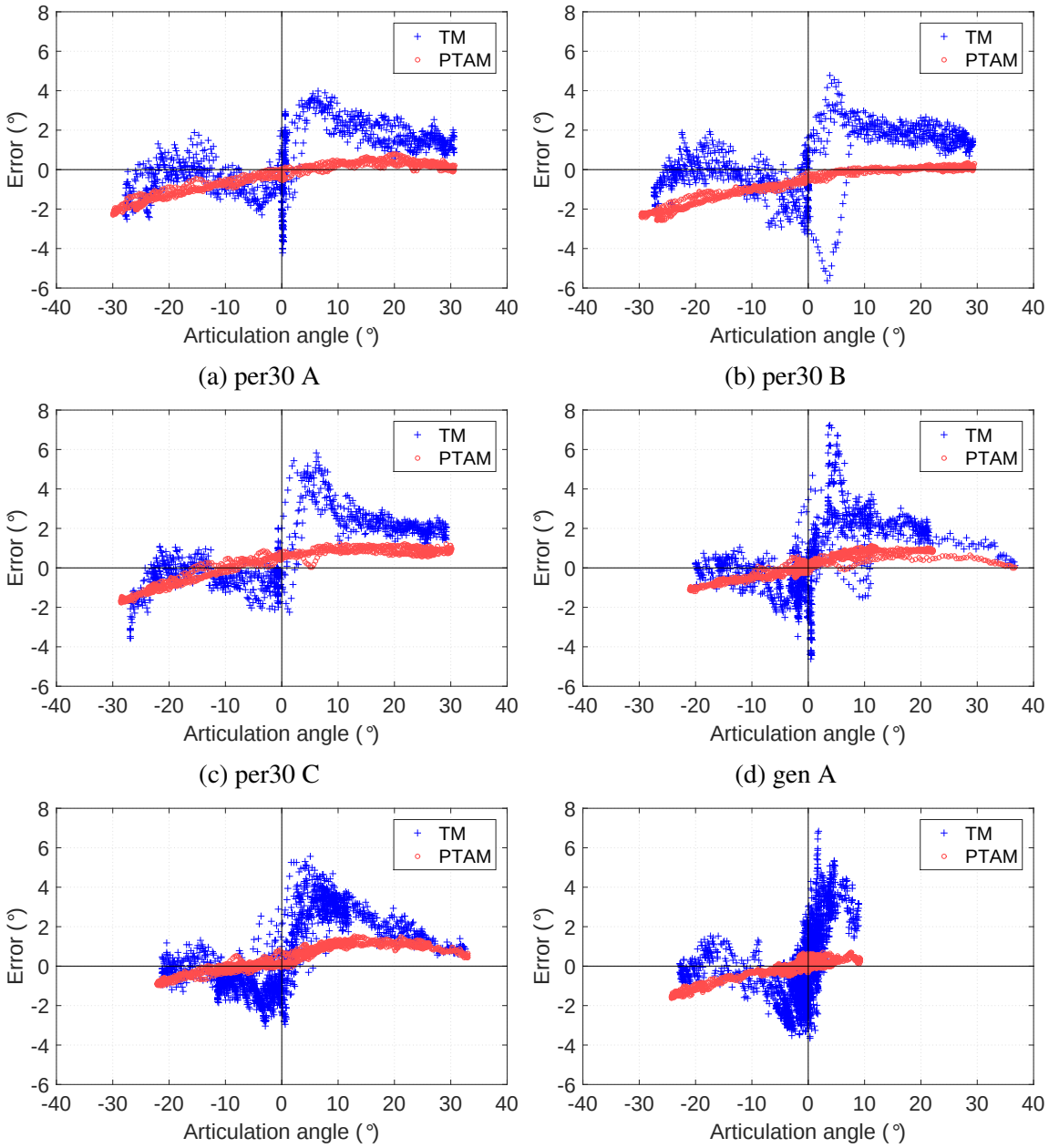


(e) Scene map (oblique view)



(f) Scene map (side view)

Fig. 4.11 PTAM frames, ‘per30 C’ vehicle test

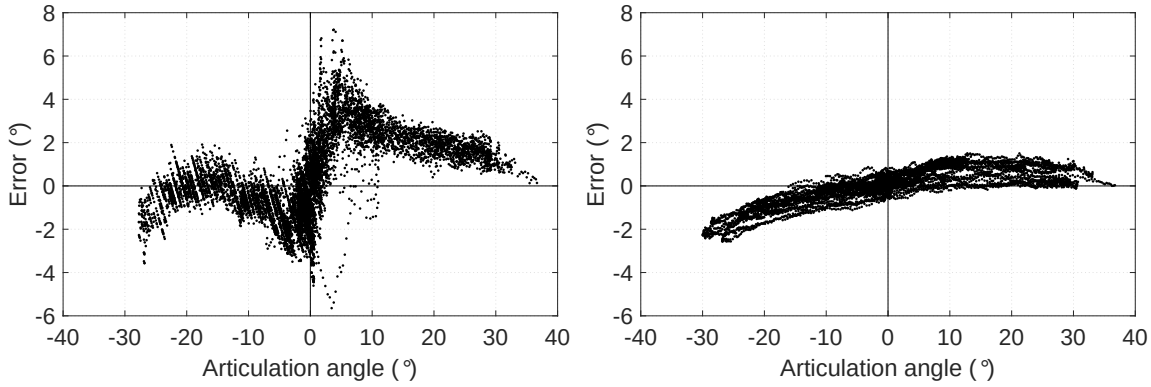


(a) per30 A (b) per30 B

(c) per30 C (d) gen A

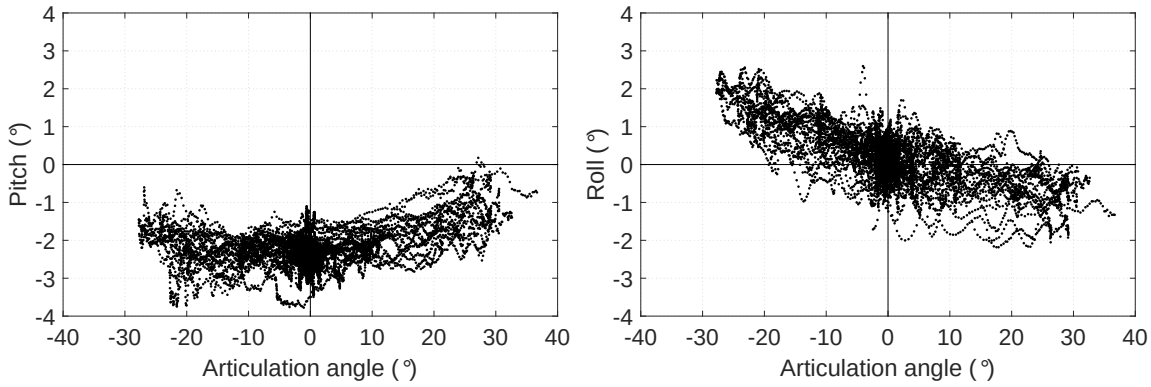
(e) gen B (f) gen C

Fig. 4.12 Errors vs. articulation angle, vehicle tests, planar trailer



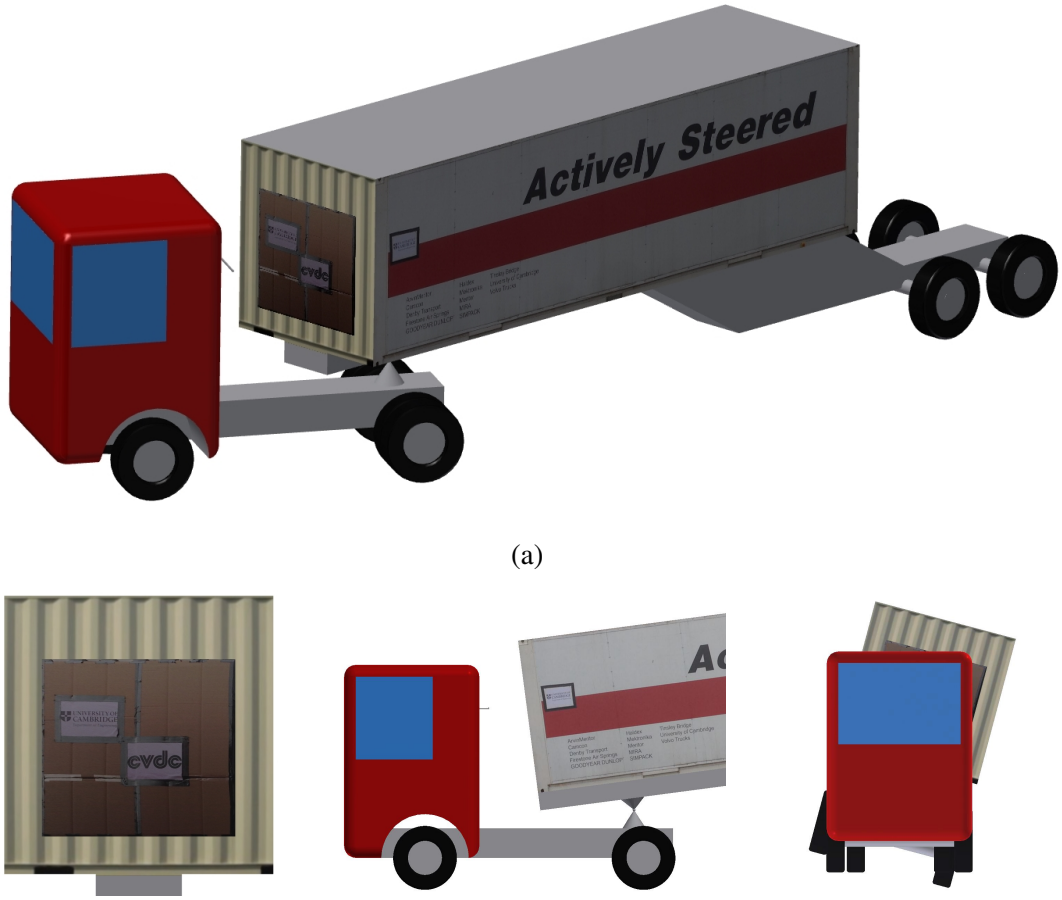
(a) TM (b) PTAM

Fig. 4.13 Errors vs. articulation angle, planar trailer, all tests overlaid



(a) Pitch (b) Roll

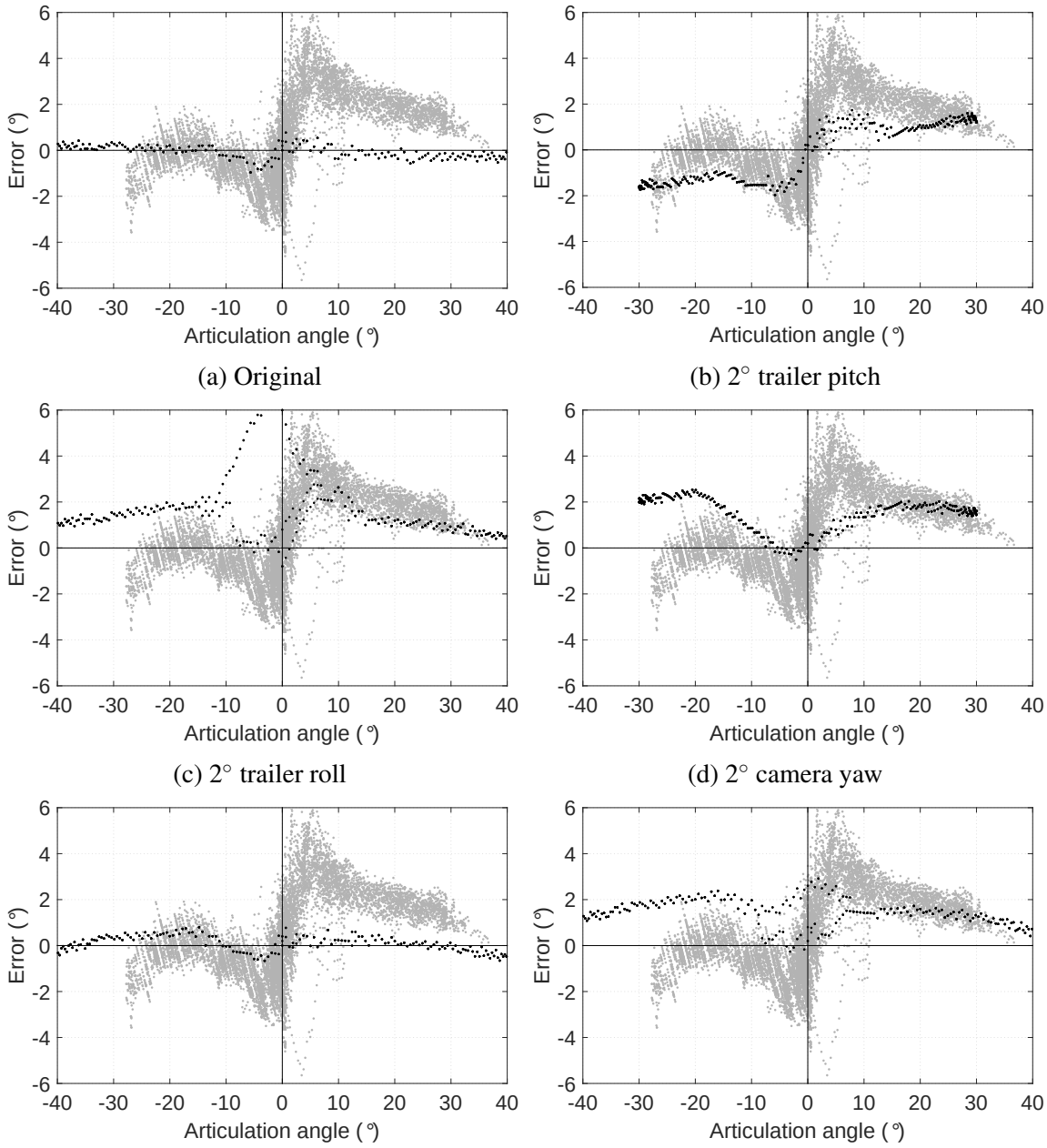
Fig. 4.14 Pitch and roll angle vs. articulation angle, planar trailer, vehicle tests, all tests overlaid



(a)

(b) (c) (d)

Fig. 4.15 CAD reproduction of test vehicle for error investigation. (a) Overview shown at 30° articulation, (b) front face detail, (c) with a large pitch angle, (d) with a large roll angle



(a) Original

(b) 2° trailer pitch

(c) 2° trailer roll

(d) 2° camera yaw

(e) 50 mm camera-hitch offset

(f) 1° trailer roll with 1° camera yaw

Fig. 4.16 Errors vs. articulation angle for the six CAD variations. Vehicle test results are shown in grey for comparison.

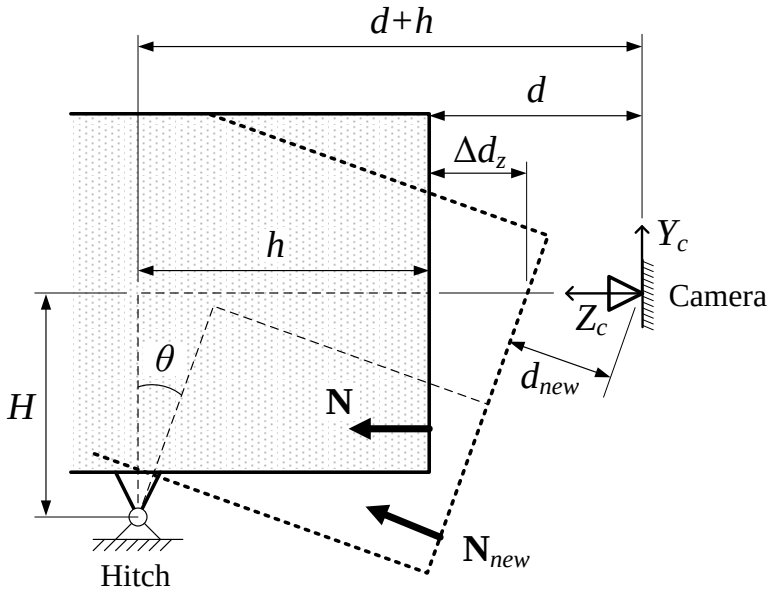


Fig. 4.17 The effect of trailer pitch angle on parameters d and $\mathbf{N}$ (side view)

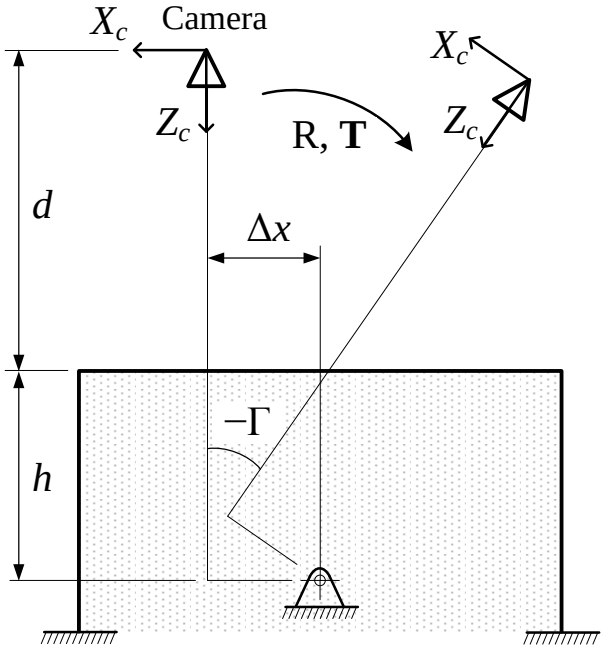
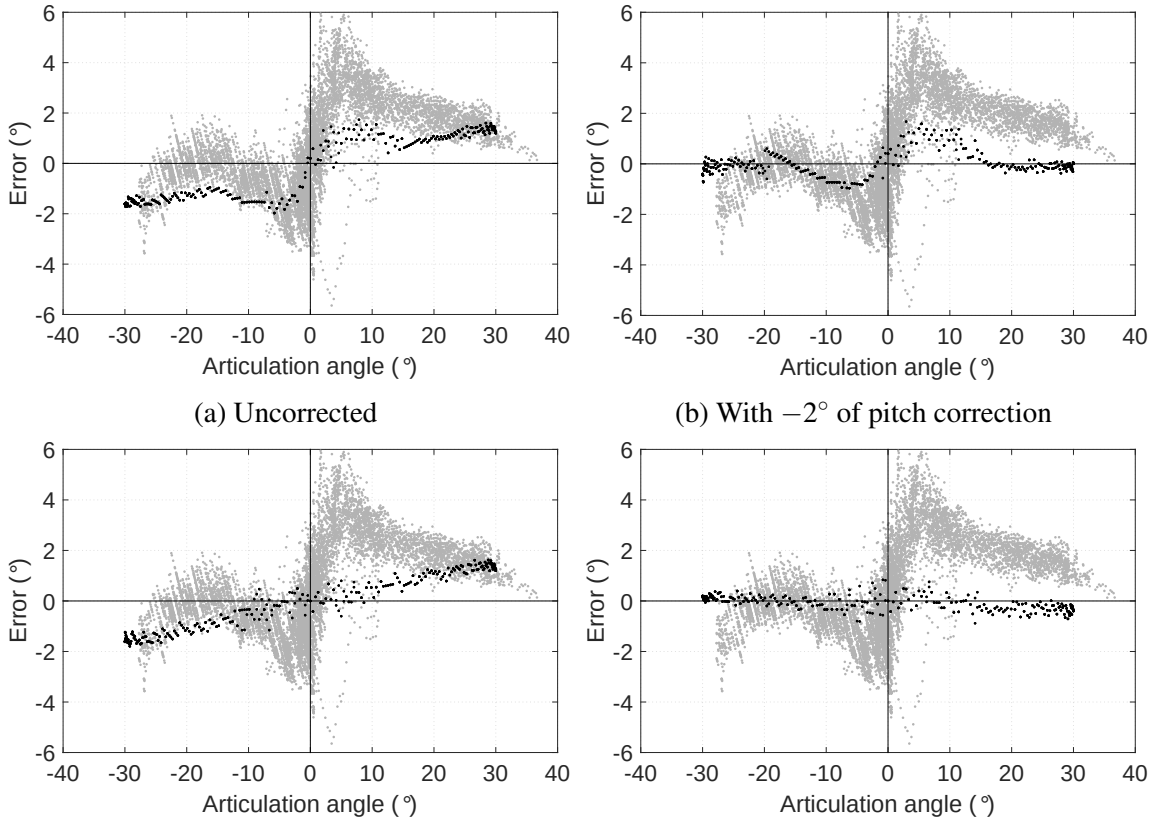


Fig. 4.18 The effect of camera-kingpin offset on relative trailer-camera motion (top view)



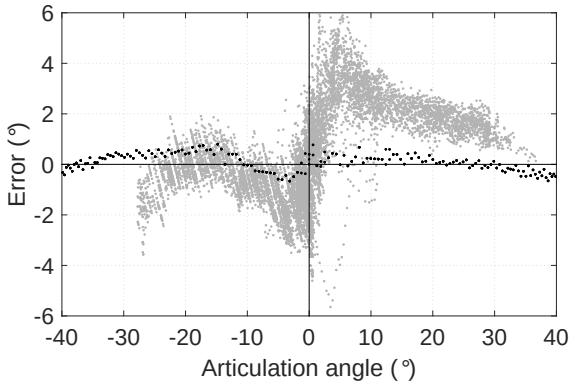
(a) Uncorrected

(b) With -2° of pitch correction

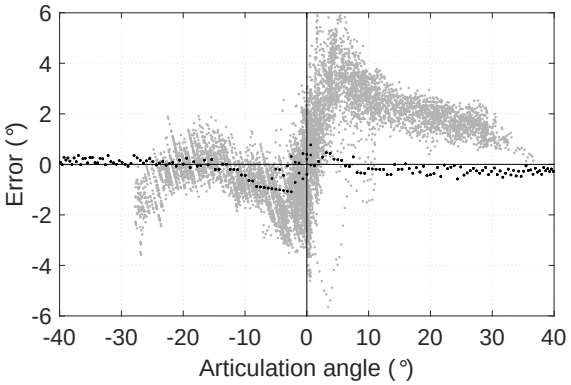
(c) With edge detection

(d) With -2° pitch correction and edge detection

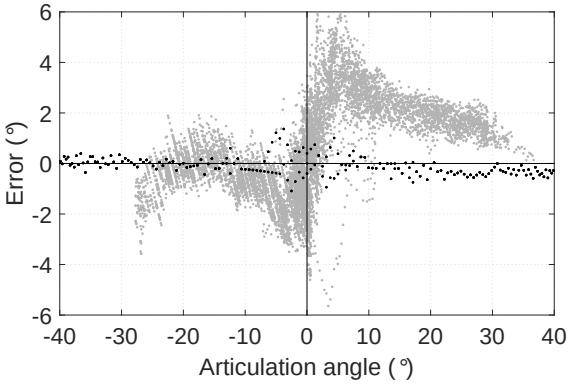
Fig. 4.19 Correction results for the 2° trailer pitch CAD scenario. Vehicle test results shown in grey for comparison.



(a) Uncorrected

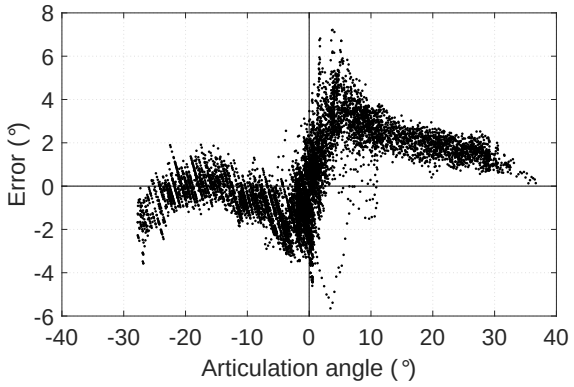


(b) With offset correction

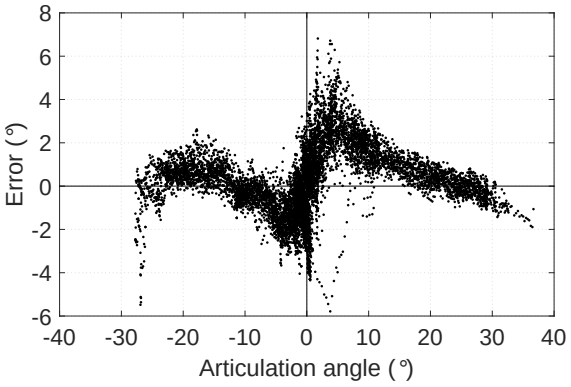


(c) With offset correction and edge detection

Fig. 4.20 Correction results for the 50 mm camera offset CAD scenario. Vehicle test results shown in grey for comparison.

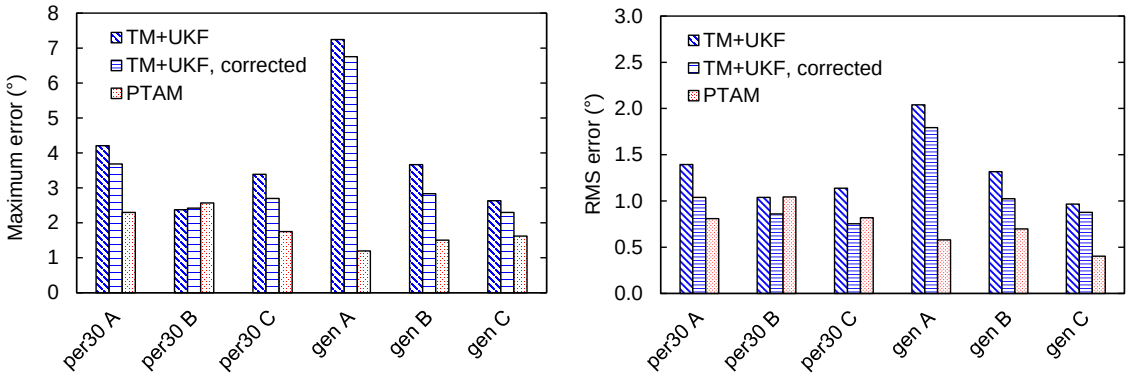


(a) No correction



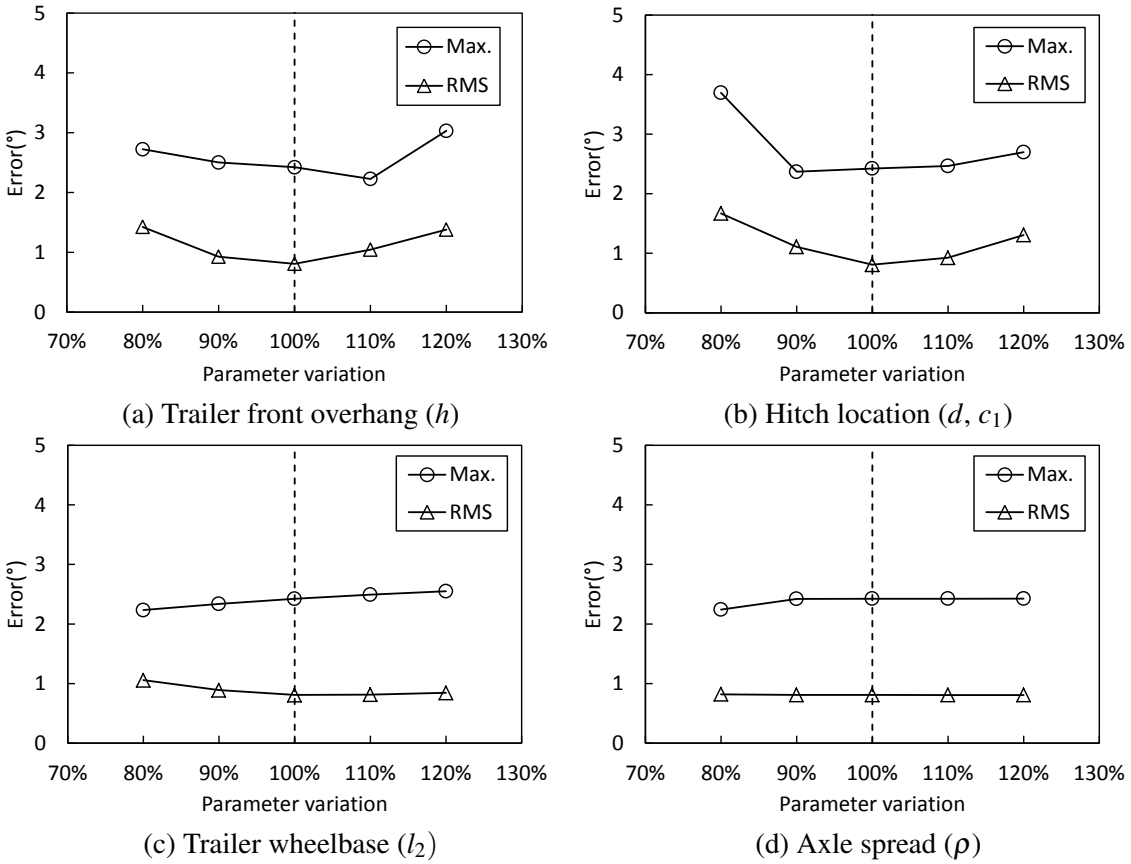
(b) With 2.2° pitch correction

Fig. 4.21 The effect of pitch correction on the error trends of vehicle tests, planar trailer, all tests overlaid



(a) Maximum errors (b) RMS errors

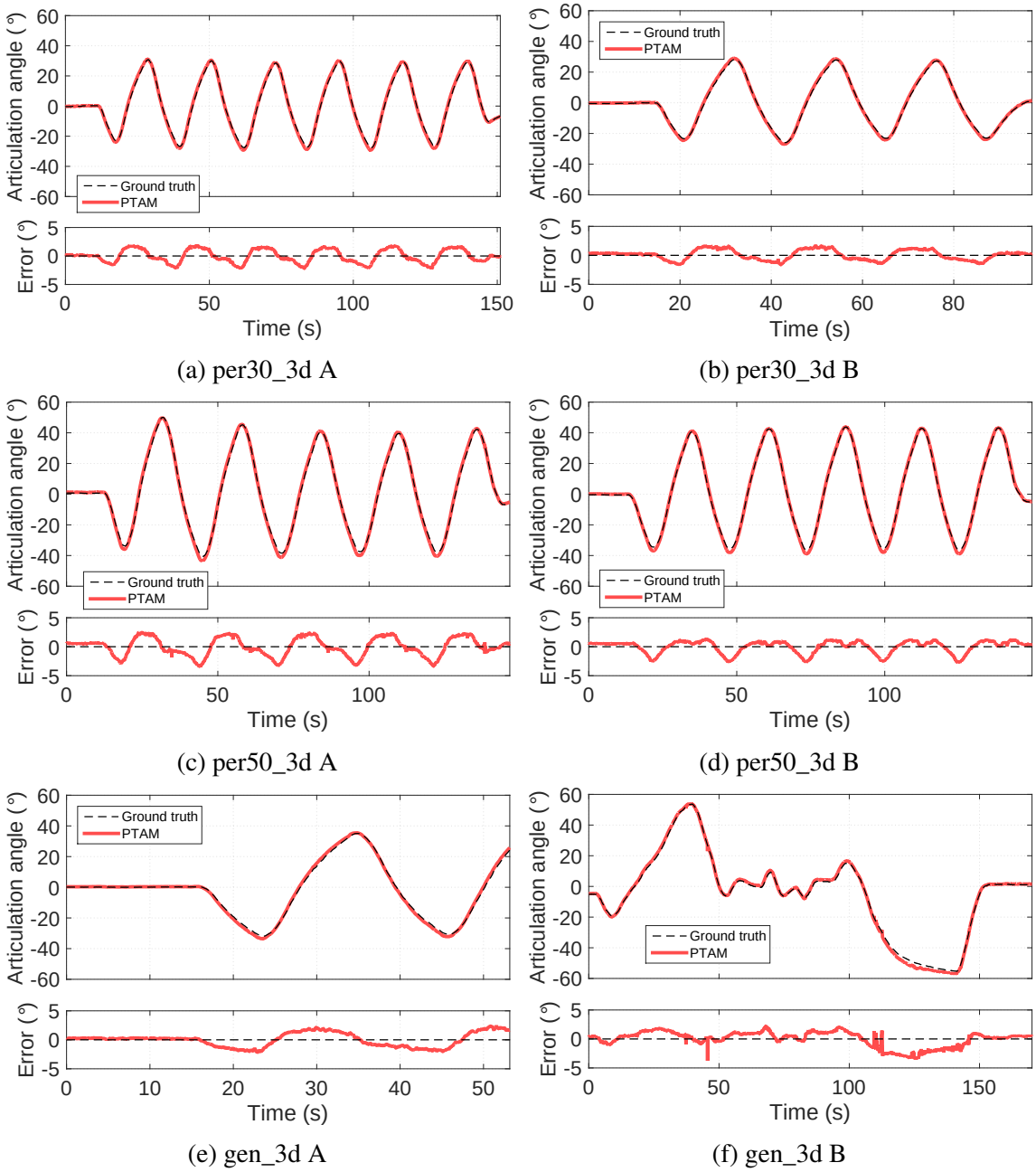
Fig. 4.22 Vehicle test errors, planar trailer front, with corrections



(a) Trailer front overhang (h) (b) Hitch location (d, c_1)

(c) Trailer wheelbase (l_2) (d) Axle spread (ρ)

Fig. 4.23 Sensitivity of the TM+UKF algorithm to variations in assumed vehicle parameters



(a) per30_3d A

(b) per30_3d B

(c) per50_3d A

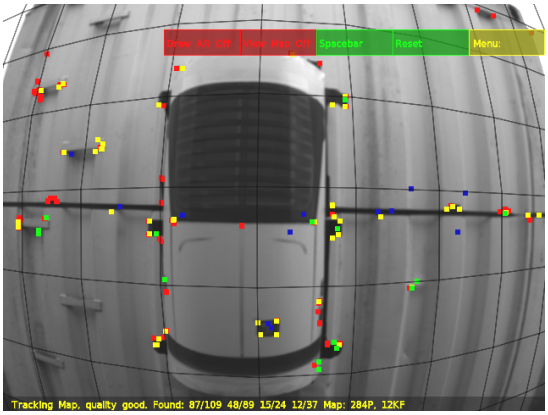
(d) per50_3d B

(e) gen_3d A

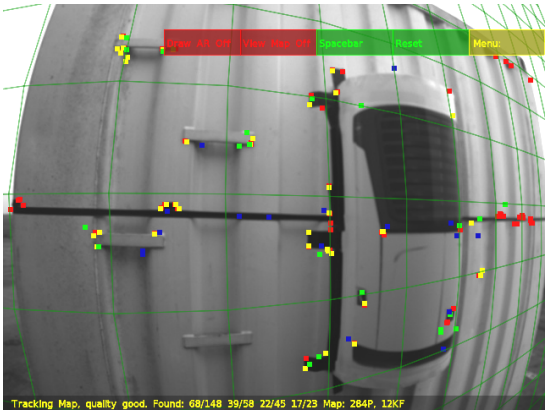
(f) gen_3d B

Fig. 4.24 Vehicle test time histories, non-planar trailer, PTAM only

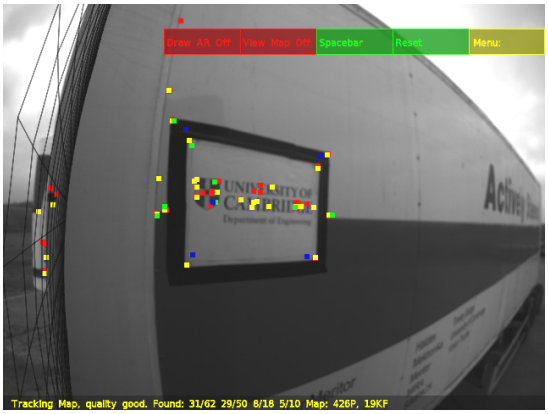
(a) Initialisation



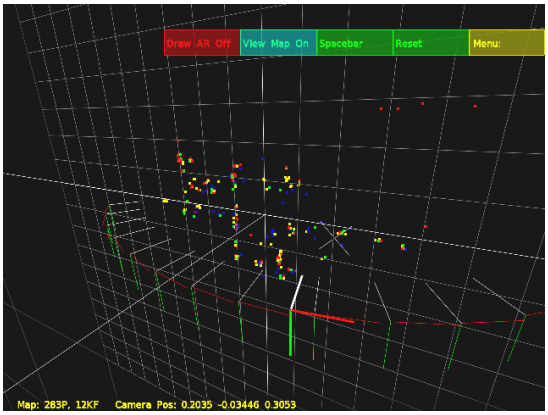
(b) Detected features



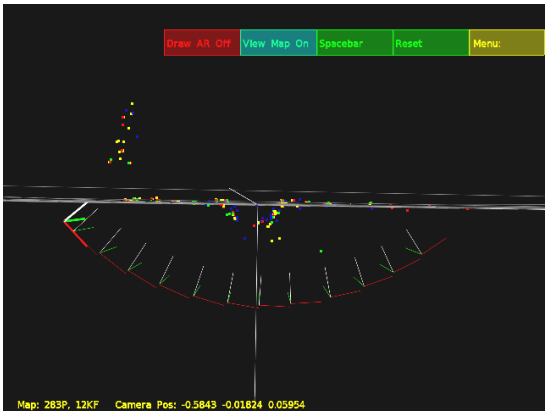
(c) Moderate Γ



(d) Very large Γ

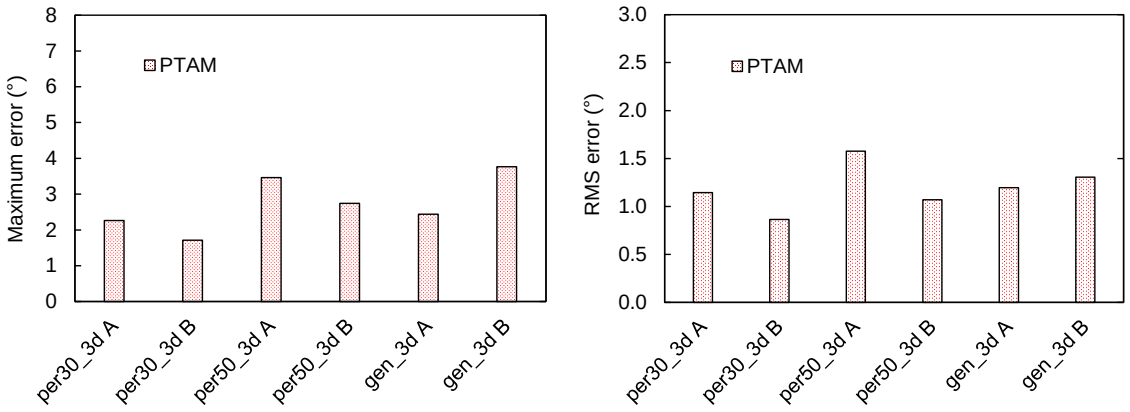


(e) Scene map (oblique view)



(f) Scene map (top view)

Fig. 4.25 PTAM details, "gen_3d B" vehicle test



(a) Maximum errors (b) RMS errors

Fig. 4.26 Vehicle test errors, non-planar trailer front, PTAM only

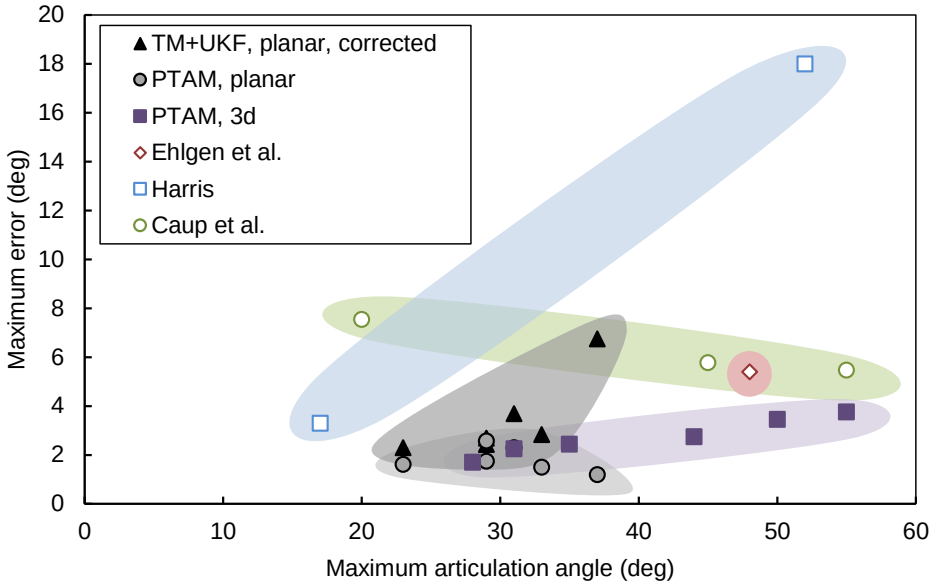


Fig. 4.27 Performance benchmark against published vehicle test data

Chapter 5

Field tests of the articulation angle sensor on a truck and full-trailer

In Chapter 3, a vision-based articulation angle sensor for articulated vehicles was developed. The concept is able to measure articulation angle at a single point of articulation, suitable for vehicle combinations such as tractor semi-trailers, and truck and centre-axle trailer combinations. In Chapter 4, the concept was validated on a full-scale tractor semi-trailer combination.

These combinations are common in Europe, and so it is important that the articulation angle sensor is compatible with them. Long Combination Vehicles (LCVs), with more than one articulation point, are increasing in use in Europe and are already common in countries including Sweden, Australia, New Zealand and South Africa. It is hence valuable to extend the applicability of the articulation sensor to include these combinations. In order for technologies such as reversing assist or jackknife control to be compatible with LCVs, articulation angle measurements at multiple points of articulation are required.

Common LCVs include 'B-double' and 'truck and full-trailer' combinations, illustrated in Figure 5.1. Extending the applicability of the articulation angle sensor in its current form to a B-double would be challenging, given that the first trailer (the 'B-link') would obscure the view of the second trailer (the semi-trailer). Alternative camera mounting options including an elevated mount above the tractor or cameras mounted to the side mirrors may be possible, but these present practical limitations and may require significant modifications to the current sensor concept. These camera-mounting options are illustrated in Figure 5.1a.

Extending the applicability of the articulation angle sensor to a truck and full-trailer combination can be achieved with relatively minor changes to the current concept. Mounting the camera at the rear of the rigid truck provides a field of view which incorporates both the

dolly and the semi-trailer without obstruction, so that the articulation angles of both can be calculated. This is illustrated in Figure 5.1b.

The truck and full-trailer combination, typically 25.25 m in length, has been widely adopted in Sweden and Finland and is known locally as the 'Nordic combination'. Although the combination is not yet widely adopted in other European countries, EU Directive 96/53 EC [39] permits any EU member state to implement LCVs including the Nordic combination. Given the combination's successful and wide-spread adoption in Scandinavia, it is likely that adoption could increase throughout Europe in the near future through Directive 96/53 EC.

In this chapter, the articulation angle sensing concept is developed further in order to measure multiple articulation angles from a single truck-mounted camera. The PTAM-based method was taken forward for this, owing to its superior performance and robustness compared to the template-matching method demonstrated in previous chapters. This chapter focusses on the implementation of the concept on a truck and full-trailer combination, however the methodology could be used to extend its use to other multiply-articulated vehicle combinations in future work.

5.1 Experimental setup

5.1.1 Vehicle and instrumentation

A Nordic combination vehicle was made available for testing by Volvo Group Trucks Technology, Sweden, and is shown in Figure 5.2. The combination consisted of a 6×2 rigid truck (with the tag axle lifted), a 2-axle converter dolly, and a 3-axle semi-trailer. The available truck was bare of any superstructure, but the addition of typical superstructures would not be expected to obscure the camera field of view. The semi-trailer had a planar front, though as discussed previously this is not a requirement.

The same Point Grey Flea3 USB camera with Fujinon lens used previously (see Chapter 4) was fixed to a length of steel square tube with brackets at each end, and attached to the truck chassis via four strong magnets (Figure 5.2, bottom right). A layer of rubber was added between the magnets and chassis to minimise the possibility of electrical grounding. The camera was positioned approximately along the longitudinal axis of the truck, and facing rearwards. From this location the field of view of the camera included both the drawbar of the dolly and the front of the semi-trailer.

'Ground truth' articulation angle measurements were obtained via a VSE sensor at the fifth wheel (as used in Chapter 4), and a custom drawbar sensor provided by Volvo at the pintle hitch (Figure 5.2, bottom left). The pintle hitch sensor consisted of a rotary

potentiometer mounted at the hitch axis, with a steel arm joining it to a bracket clamped to the drawbar. Measurements from the VSE sensor had a resolution of 0.05° as before. The drawbar sensor had a comparable resolution, but an estimated accuracy of approximately 0.5° (provided by Volvo Group Trucks). Gains for the two sensors were pre-calibrated by Volvo. Offsets were obtained on the day of testing during straight-line driving along the oval track at the test site (see next section).

Although previous results showed that artificial markers were not necessary for the PTAM-based concept, some visual texture was added to the semi-trailer and drawbar for these tests. Additional tests without the markers were conducted to evaluate sensitivity to this. The semi-trailer with and without markers, together with the marked drawbar, are shown in Figure 5.2, top.

From the point of view of the OEM (*i.e.* Volvo in this case), it is important for ancillary equipment such as a reversing assist camera to be mounted below the top of the main I-beams of the truck chassis. This allows for a variety of superstructures to be fitted. The test truck was already fitted with an analogue reversing support camera, which was mounted in what would be an ideal location for the articulation sensing camera (shown bottom right in Figure 5.2). The location of the digital camera was sufficiently near to the existing analogue camera that the fields of view were assumed to be comparable.

A summary of the instrumentation and communications layout is illustrated in Figure 5.3. The camera was connected via USB 3.0 to a laptop computer in the truck cab. The camera was calibrated using the PTAM calibration module described previously, using the same 7×10 checkerboard. Calibration results are given in Appendix C. Analogue signals from the two articulation angle sensors were digitised and logged using a dSPACE MicroAutoBox® in the cab, provided by Volvo.

Synchronisation signals were sent from the Autobox via CANbus to the laptop. On the laptop a Python script was used to capture images and log sensor data at 20 Hz. Images were captured using the FlyCapture and OpenCV libraries for Python, while the PCANBasic library was used to interpret CAN messages from the Autobox.

Processing of image sequences in PTAM was performed offline on the same 2.4 GHz laptop computer as before (and the same laptop computer used to log data here). Detailed specifications are provided in Appendix A.

5.1.2 Test scenarios

Testing was carried out in September 2016 at Hällered, Volvo's vehicle proving ground near Gothenburg, Sweden. The trailer storage yard next to the mechanical workshop was chosen to perform tests. This provided the visual scenery and manoeuvres most representative of

truck loading and unloading areas, where reversing assist systems would be most applicable. Straight sections of the oval track at Hällered were used to obtain the offsets of the reference articulation angle sensors before testing.

Selected low-speed manoeuvres were carried out in the yard, around parked trailers and around the workshop building. The testing site and manoeuvres are illustrated in Figure 5.4. Six manoeuvres were carried out:

1. a left turn (anti-clockwise) loop around the central trailer parking area ('left turn'),
2. a right turn (clockwise) loop around the central trailer parking area ('right turn'),
3. an arbitrary driving route around the workshop ('arbitrary A'),
4. a shorter arbitrary driving route around the workshop ('arbitrary B'),
5. a short duration manoeuvre over an uneven unpaved surface ('uneven'), and
6. the 'left turn' manoeuvre, with semi-trailer markers removed ('no markers').

The left turn manoeuvre was an easily repeatable manoeuvre, and so was chosen for the no markers test to offer a good comparison of performance. The right turn manoeuvre would have equally sufficed. Three runs of each manoeuvre were performed.

Each test was started with the vehicle stationary and oriented such that both articulation angles were non-zero. This ensured that when the vehicle started moving, lateral motion of both the semi-trailer and drawbar relative to the camera would occur, as was necessary for the stereo initialisation process[1]. In some cases the motion was found not to be sufficient, and so the initialisation was performed later in the manoeuvre.

Each manoeuvre included a section where the vehicle was travelling straight, so as to zero the co-ordinate frame. The stereo initialisation and zeroing processes are discussed in more detail in the following section.

5.2 PTAM implementation

Sample images from the camera are shown in Figure 5.5 at both zero and non-zero articulation angles. For the purposes of extracting the two independent articulation angles, the images

[1]Recall from Chapter 3 that the PTAM stereo initialisation process involves taking two images of the scene from two viewpoints separated by some relative translation. These are the first two 'keyframes' which are used to define the initial 3-D map of the scene. Because no depth information is known (from a stereo camera for example), the 3-D map is only known up to an arbitrary scale factor. Additional keyframes can be added over time as new features are detected. In this case the camera is fixed to the truck, the reference body, and the scene is moving relative to the camera. Therefore translation of the trailer or drawbar relative to the camera is required in order to capture the two keyframes for stereo initialisation.

were partitioned into two regions of interest: semi-trailer and drawbar. These partitions are shown in the figure. The sizes of the partitions were 440×270 and 560×180 pixels for the semi-trailer and drawbar respectively, and were centred laterally.

The precise dimensions of these regions were chosen by trial and error to give a good field of view of the target object within the range of articulation, whilst minimising unnecessary background visuals. It is likely that an optimum size of these partitions could be found which would be suitable for most or all anticipated trailers and geometries; this was not investigated in this work.

The two partitions were processed separately by applying suitable masks over the image outside of the region of interest. It is envisaged that in future work suitable modifications could be made to the PTAM algorithm to process both regions simultaneously. This could be achieved by adding another two processing threads (a tracking and a mapping thread for the second partition). Alternatively, the two regions could be processed sequentially at each frame. The small lag caused might be acceptable, otherwise suitable delay compensation techniques could be adopted.

The total number of features tracked in both semi-trailer and drawbar regions of interest are likely to be equal to or less than when the full frame was used in the semi-trailer tests of Chapter 4. For the same number of keyframes, the processing demand varies proportionately with the number of features (see [123]). As such, with the same number or fewer features, the total processing demand in a sequential processing routine should yield comparable framerates to the full-frame tests.

From Figure 5.5 the view of the semi-trailer can be seen to be similar to that of the tractor semi-trailer tests (see Figure 4.3). However, the semi-trailer is further away from the camera, and given the added articulation point we can expect more lateral translational motion. Any effect of these differences should be evident in the results.

The drawbar view is slightly different in that the dominant plane is horizontal instead of vertical. However, this was found not to be of significant importance. What is important about the drawbar is the proximity of the camera to the point of rotation. If the camera was located on the axis of rotation stereo initialisation would not be possible, as this requires relative *translation* between the two initial keyframes.

The location of the camera was sufficiently offset from the axis of rotation to ensure that stereo initialisation of the drawbar was possible in all cases. However, relatively large rotations of the drawbar were required between the two keyframes to ensure an adequate initialisation. In future work, an increased offset position of the camera might help reduce the initialisation articulation angle. The stereo initialisation procedure was otherwise identical to Chapter 4.

An additional initialisation step was introduced to zero the articulation measurements during processing, in a manner which could be easily carried out during real-time processing. After stereo initialisation, the reference co-ordinate frame is not necessarily aligned with the longitudinal axis or yaw plane of the truck, depending on the features found or the orientation of the trailer between the first two keyframes.

When the trailer was known to be travelling straight, with both articulation angles approximately zero, a command was given in the C++ script which obtained the instantaneous rotation matrix of the pose. Subsequent rotation matrices were post-multiplied by this reference matrix to adjust the co-ordinate frame accordingly, before performing Euler decomposition to extract the articulation angle.

An example of the stereo initialisation and zeroing procedure is given in Figure 5.6, showing a time history of PTAM and ground truth measurements. When some lateral movement of the semi-trailer or drawbar begins, the first keyframe is obtained. Once sufficient additional lateral motion has occurred, the second keyframe is obtained at approximately 12 seconds, defining the initial stereo map. Hereafter the co-ordinate system is not necessarily aligned with the vehicle, as is evident in the measurement bias. At approximately 72 seconds, it was observed that the trailer was straight, and so the zeroing command was given, applying a zeroing correction to subsequent measurements.

In processing, every run was processed once in full, with stereo initialisation and zeroing as described. However, to obtain meaningful error statistics over the full duration of the run, each run was re-processed using the initial stereo map and zeroing correction, giving initialised and zeroed articulation measurements from $t = 0$ seconds.

Processing time was 15 ms for the trailer and 13 ms for the drawbar including distortion correction and stereo rectification. The total time is comparable to the original PTAM algorithm, and permits frame rates of up to 20 fps.

5.3 Results

Example stereo initialisation keyframes and subsequent initialised feature maps are given in Figure 5.7. The marked semi-trailer (a), unmarked semi-trailer (b), and drawbar (c) are shown. Recall that the graphical planar grid superimposed on the trailer represents the dominant plane found in the scene, based on the locations of detected features (intended for Augmented Reality purposes in the original PTAM paper [123], but retained here for visual feedback).

It is clear that the markers provided on the semi-trailer (Figure 5.7a) introduced additional feature points compared with the unmarked semi-trailer (Figure 5.7b), but that the unmarked

semi-trailer still provided sufficient features for initialisation and tracking. In both cases there was a concentration of features around the service connections and along the bottom edge of the semi-trailer around rivets, trailer name plates and trailer identification code (indicated in Figure 5.7b).

These features are likely to exist on almost all trailers as a bare minimum, providing a good lateral distribution of features which is necessary for yaw rotation measurement. This is an important finding, and confirms that sufficient features may be found and tracked without the need for any markers.

In the case of the drawbar (Figure 5.7c), it is clear that a number of features were obtained from existing points such as service connectors, name plate, bolt heads and structural discontinuities. The added markers contributed a number of additional features, but it is assumed that only the base features would be sufficient given the observations with the unmarked semi-trailer.

It is also clear that a larger rotation was required during the drawbar stereo initialisation process. If the motion was too small, the initialisation would either fail or yield obviously erroneous and/or erratic measurements. In practice, it would be straight-forward to detect a failed initialisation, or to otherwise ensure sufficient motion between keyframes in an automatic fashion. Adding a lateral offset between the camera and hitch axis would help reduce the angle required for initialisation, by increasing the translation component of the relative motion ($\mathbf{T}$).

Time histories of articulation angle measurements for the six manoeuvres are shown in Figure 5.8 (drawbar) and Figure 5.9 (semi-trailer). PTAM and ground truth measurements are plotted, and the error signal between the two is shown beneath each plot. Results for run 1 (of 3) from each manoeuvre are plotted. PTAM measures total relative pose, and so in the case of the semi-trailer this is the *sum* of drawbar and semi-trailer articulation angles. Therefore, in this case the sum of pintle hitch and VSE sensor measurements were used as the ground truth.

Measurements from the pintle hitch sensor were unfiltered and exhibited a noise component. The total ground truth articulation angle measurement between truck and semi-trailer was hence also noisy, although the fifth wheel sensor measurements were not. To address this, the pintle hitch articulation measurements were filtered in post processing using a second-order low-pass Butterworth filter at a cut-off frequency of 0.1π rad/sample (2 Hz in this case). The filtered signals were used to calculate the error signals.

The trends of the PTAM and ground truth measurements show good agreement, and errors are mostly small with smooth variations with articulation angle. There is little evidence of bias except for the scenario of the trailer on uneven ground (Figure 5.9e).

Bias will occur if the system is inaccurately zeroed, and so it is important that this step is performed correctly. This was difficult to perform accurately for the uneven ground scenarios as there was no consistent region of zero articulation. This was a limitation of the manoeuvring area used for these tests (see Figure 5.4). Small variations in initialisation and zeroing accuracy can be expected in the other manoeuvres too due to the manual nature in which they were performed, however it is likely that this process could be better automated in future.

The manoeuvres conducted provided a large variation in background scenery, and the results show no discernible sensitivity to this. Results with and without markers are very similar (Figure 5.8a vs. Figure 5.8f, Figure 5.9a vs. Figure 5.9f), despite the lack of artificial markers.

It is clear from Figures 5.8 and 5.9 that the PTAM measurements consistently underestimate articulation angle compared to the ground truth. This is comparable to the observations in Chapter 3 (Figure 3.23) but contrary to Chapter 4 where angles were overestimated relative to the ground truth (Figure 4.13b). However, in all three cases an approximately linear sensitivity was observed. Errors as a function of articulation angle for each of the manoeuvres are shown in Figures 5.10 and 5.11, from which the consistency and approximate linearity of the variations are clear.

A summary of RMS and maximum errors per manoeuvre is given in Figure 5.12. For each manoeuvre, data for each of the three runs as well as the average result per manoeuvre are presented. The larger errors for the semi-trailer can be attributed to respectively larger articulation angles. The similarity in RMS and maximum errors between the 'left turn' and 'no markers' scenarios is clear for all three runs. The largest RMS and maximum errors were obtained during the manoeuvres on the uneven road surface, and these results are distinctly outlying compared to the other manoeuvres. All other manoeuvres yielded comparable results.

The original purpose of the manoeuvres on the uneven road surface was to assess PTAM's robustness to erratic roll motion caused by undulating road surfaces. These manoeuvres yielded significant periodic roll motion over the extent of the manoeuvre, and yet there is no evidence of periodic variation in errors (the noise component is comparable to the other manoeuvres). Therefore, these errors must originate from a different source.

One possible source is the difficulty with which these manoeuvres were zeroed, and this is evident in Figure 5.9e. Secondly, there appears to be a consistent 'spike' in the errors at around the same point in the manoeuvre in all cases, without an associated increase in articulation angle (see Figure 5.9e at 60 seconds). From the image sequences it was observed that this is the point when the truck transitions from the uneven road surface to the paved

surface, resulting in a noticeable change in pitch angle between truck and trailer. This change is more significant than the transient pitch variations experienced in other manoeuvres, and appears to have a direct effect on the errors. In future work it is expected that this pitch angle effect could be accounted for with corrections to the pose matrix R before calculating the articulation angle.

Over all manoeuvres, using the three-run average results, RMS errors were in the range 0.79–0.93° (1.83°) for the drawbar and 1.07–1.51° (2.35°) for the semi-trailer. The values given in parenthesis represent the outlying uneven road surface results. Maximum errors were in the range 2.50–3.05° (4.60°) for the drawbar and 3.44–4.04° (5.05°) for the semi-trailer.

5.3.1 Error trends, Chapters 3, 4, 5

The approximately linear variation in errors with articulation angle observed in the drawbar and trailer results for the Nordic combination was also observed in the results of Chapter 3 (tractor semi-trailer, CAD simulations) and Chapter 4 (tractor semi-trailer, field tests). Figure 5.13 shows the maximum error as a function of maximum articulation angle per manoeuvre for the drawbar and semi-trailer of the Nordic combination, together with the tractor semi-trailer results of Chapters 3 and 4. The sign of the maximum errors has been included to differentiate between overestimates and underestimates of articulation angle. A linear least squares fit has been included for each case. (A zero-intercept has been forced, consistent with the behaviour observed in Figures 5.10 and 5.11.)

These results prompt some key questions:

1. Why do the errors increase approximately linearly with articulation angle?
2. What are the factors governing the magnitude and sign of these errors?
3. What are the differentiating factors between the test cases of Chapters 3, 4 and 5 which have given rise to the differences in the magnitude and sign of the errors?

The first question can be addressed by considering the how PTAM estimates pose through feature tracking. The primary set of features is captured during the stereo initialisation process, contained in the first two ‘keyframes’. These two initial keyframes contain the only set of features tracked, except for the few cases when larger articulation angles allowed keyframes from the trailer sides to be added (Chapter 4, Figure 4.25d).

Estimating the current camera pose requires detecting the original keyframe features in the current frame. To do this, the original features first undergo an affine warp (a simplified version of Equation 3.17) to account for the change in viewpoint from the original keyframe to the current frame. As the difference in viewpoint increases with increasing articulation

angle, the similarity between the estimated warped feature and the actual observed feature will have deteriorated, and so will the ability the accurately locate the new position of that feature. The template-matching algorithm is a large-scale version of this same procedure.

This is further exacerbated when the feature points are not planar. The assumption that features are locally planar (Section 3.4.2) is accurate for many of the features detected in the current vehicle tests. The assumption is also generally accurate for non-planar features, provided changes in viewpoint are small. However, the vehicle tests conducted in this work include both non-planar elements and large changes in viewpoint. Non-planar elements range from small discontinuities at welded joints (see Figure 4.25c, features in the left half of the image) to highly non-planar elements such as the coiled service cables (Figure 5.7b).

It is therefore reasonable to expect that errors should increase approximately linearly with increasing articulation angle relative to $\Gamma \approx 0^\circ$ (where the initial keyframe was obtained during initialisation). For the cases where additional keyframes from the trailer sides are added, a sudden reduction in errors is not expected as the prior pose estimate for the new keyframe is based on the existing estimates, which are based on the original keyframes.

If the image sequences were significantly longer in duration than investigated here, the effects of global bundle adjustment could become evident (see Section 3.4.2), and should help reduce errors. The addition of more keyframes from the trailer sides may also help to improve errors over time during global bundle adjustment.

If this is the primary source of the observed error trend, then it must be established why these errors should manifest themselves differently between the tests conducted in Chapters 3, 4, 5. From Figure 5.13 it is clear that neither the gradient of the linear trend nor the sign of the errors (overestimating or underestimating) is consistent.

The reason for this could be a one of multiple factors, including camera location relative to the trailer, the vertical distribution of features on the trailer face or camera calibration. These factors were different for each set of tests, and so may help to explain the differences.

In Chapter 3, the camera was slightly below the centre of the trailer face, though features were well distributed vertically (Figure 3.24b). In Chapter 4, the cameras were slightly above the centre of the trailer, with a similarly good vertical distribution of features on the trailer face, and a slight prominence towards the top half of the trailer (Figures 4.11c, 4.25c). In the current chapter, the camera was mounted significantly lower than the centre of the trailer face, and as a result features were predominantly in the lower portion of the trailer face. These differences in camera position and feature distribution are consistent with the differences in the sign of the errors.

In the vehicle tests, it is possible that small camera calibration errors could have resulted in the increasing errors with articulation angle. For the case of the tractor semi-trailer and

the drawbar of the Nordic combination, increasing articulation angle is directly correlated with movement of the trailer face away from the centre of the image. For the semi-trailer of the Nordic combination this is not always true given the two articulation points, but it is true during much of the cornering manoeuvres conducted during these tests.

Errors in either focal length or the radial distortion coefficient (see Equation 2.11) would therefore result in increasing errors with articulation angle, over- or underestimating articulation angles consistently for both $+\Gamma$ and $-\Gamma$ as was observed. Errors in the principal point (especially u_0) would more likely give rise to overestimates in one direction and underestimates in the other, which was not observed.

External factors such as a pitch angle offset also play a part in the observed error trends, as was proposed to explain the high errors in the uneven road surface manoeuvres. In general, this could be a result of either trailer pitch or pitch in the camera mounting. Alternatively, it could result from the initial orientation of the reference plane from which PTAM takes pose estimates. This reference plane is fitted to the dominant plane in the scene during initialisation, which in all tests has been the majority of the trailer face. However, some slightly out-of-plane features (such as on the services connectors) may have resulted in small pitch offsets of the plane relative to the trailer.

It is possible that pitch angle offsets could be accounted for using the pitch angle measurement from PTAM's rotation matrix, but this will be sensitive to the initial orientation of the dominant plane observed during initialisation. Alternatively, the correct alignment axes of the trailer could be determined during some additional initialisation manoeuvring, by determining the vertical axis of trailer articulation. Subsequently measured variations in pitch angle could then be used to correct for the effect of pitch on articulation angle estimates. An investigation into pitch angle correction and feature distribution effects is recommended for future work.

5.3.2 Comparison with published results

Figure 5.14 shows an updated version of Figure 4.27, where the Nordic combination vehicle test results have been compared to published results, as well as earlier results from this work. Recall that maximum errors are used in the comparison as RMS errors were not available from the literature. The three-run average result for each manoeuvre is plotted for simplicity.

The new results sit well with the previous PTAM results (with the exception of the uneven road surface results), and further reinforce the performance benefits over published sensing solutions and the template-matching approach. The Nordic combination tests also provide results at the highest articulation angles assessed, 68° (arbitrary A, run 3).

5.4 Conclusions

1. The PTAM-based articulation angle sensing concept was modified to be compatible with a Nordic combination vehicle (or truck and full-trailer), and to independently measure the articulation angles of both a drawbar and semi-trailer with one camera.

2. Vehicle tests were conducted on a Nordic combination at Hällered, Sweden. Testing was conducted in a trailer parking and workshop area, giving a good visual representation of a trailer docking yard where such as system might be utilised.

3. Six low-speed manoeuvres were conducted with articulation angles of up to 68°. Manoeuvres included typical low-speed manoeuvring, driving over uneven road surfaces, and tests with and without markers on the semi-trailer.

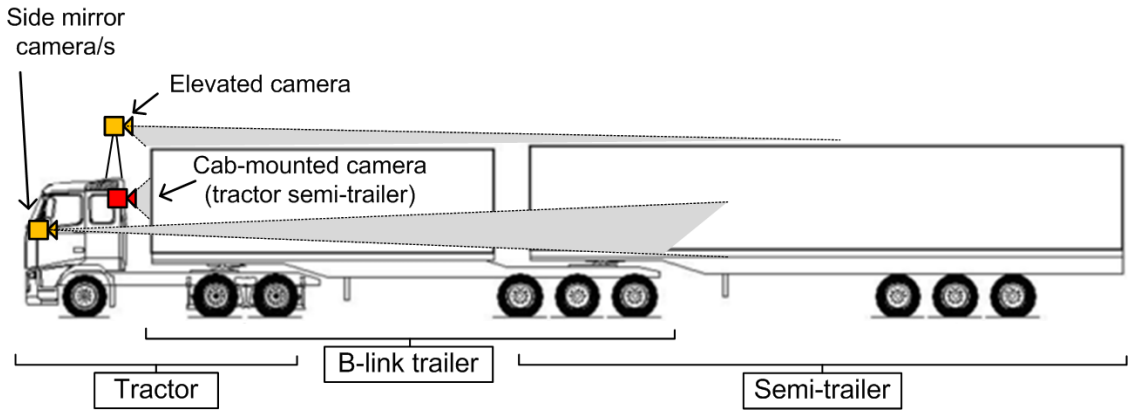
4. A new semi-automated zeroing process was introduced which should be straightforward to perform automatically in future work.

5. RMS errors were in the range 0.79–1.83° for the drawbar and 1.07–2.35° for the semi-trailer. Maximum errors were in the range 2.50–4.60° for the drawbar and 3.44–5.05° for the semi-trailer. No differences in error were observed for comparable tests with and without artificial markers. Manoeuvres on uneven road surfaces yielded large errors inconsistent with the other manoeuvres. It was proposed that this was the result of a combination of zeroing accuracy and large changes in truck-trailer relative pitch angle during the manoeuvre.

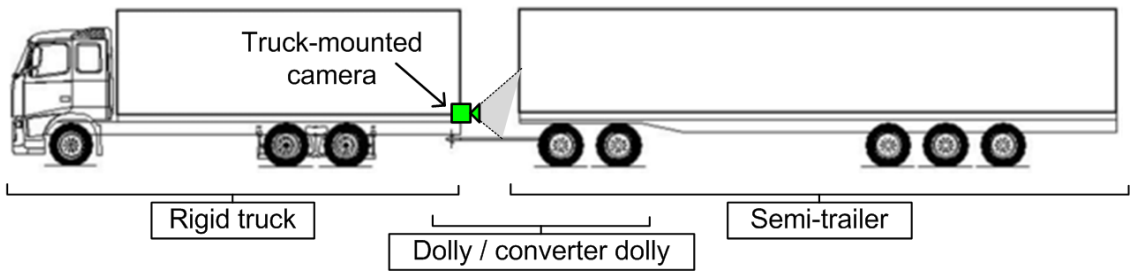
6. Errors were shown to increase smoothly as a function of articulation angle, consistently underestimating the angle relative to the ground truth. The increase in errors with articulation angle was consistent with results observed in Chapters 3 and 4, but variations in the magnitude and sign of the errors between these sets of results were not consistent.

7. Increasing errors with increasing articulation angle are expected given PTAM's inherent feature-based method for pose estimation. Factors which could give rise to differences in the magnitude and sign of errors between Chapters 3, 4 and 5 include vertical camera location, vertical distribution of feature points, camera calibration, and pitch angle offsets. An investigation into these effects is suggested for future work.

5.5 Figures



(a) B-double



(b) Truck and full-trailer combination (or Nordic combination)

Fig. 5.1 Camera-mounting options for multiply articulated vehicles (vehicle illustrations from [1])

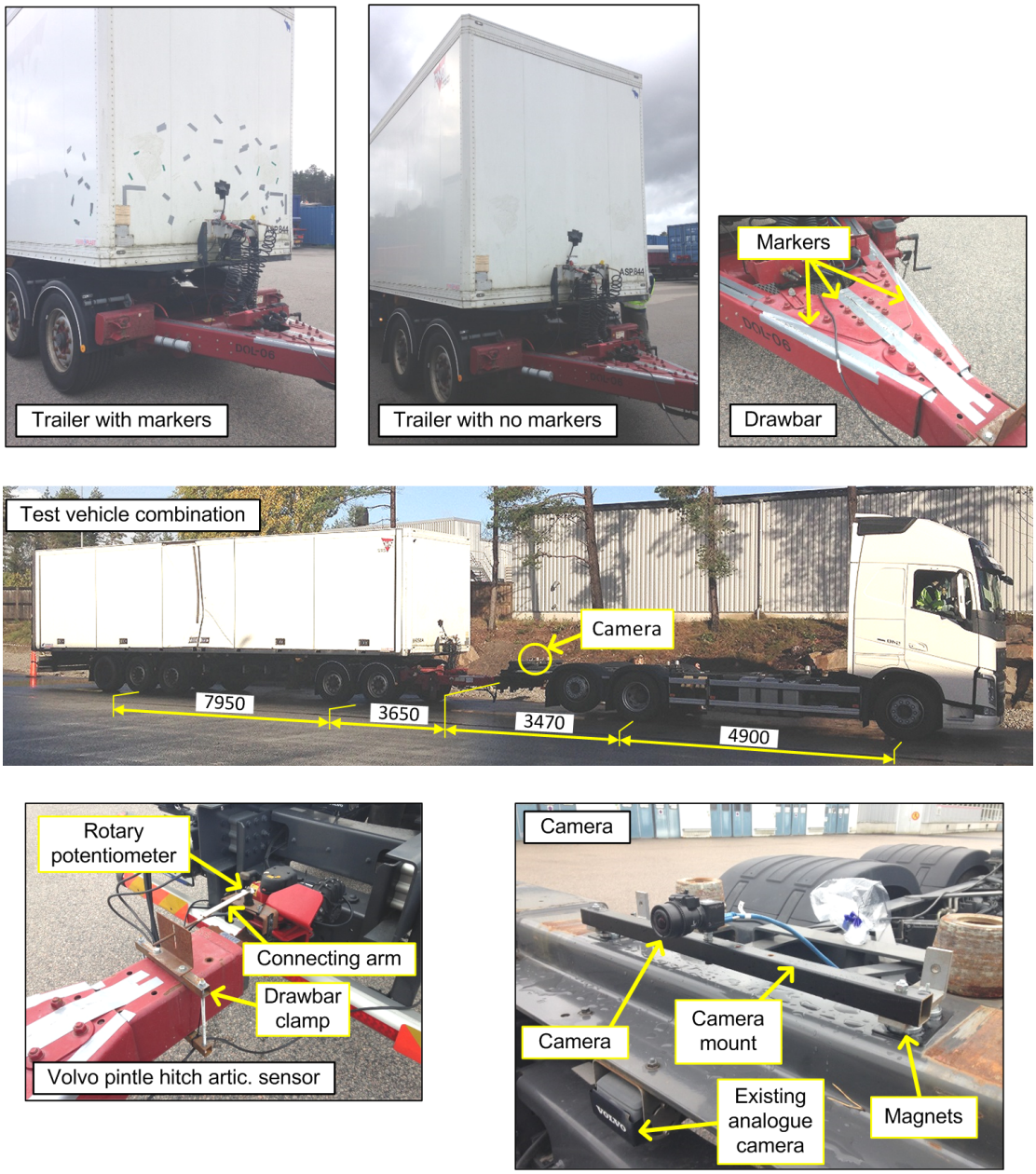


Fig. 5.2 Nordic combination test vehicle and detail (dimensions in mm)

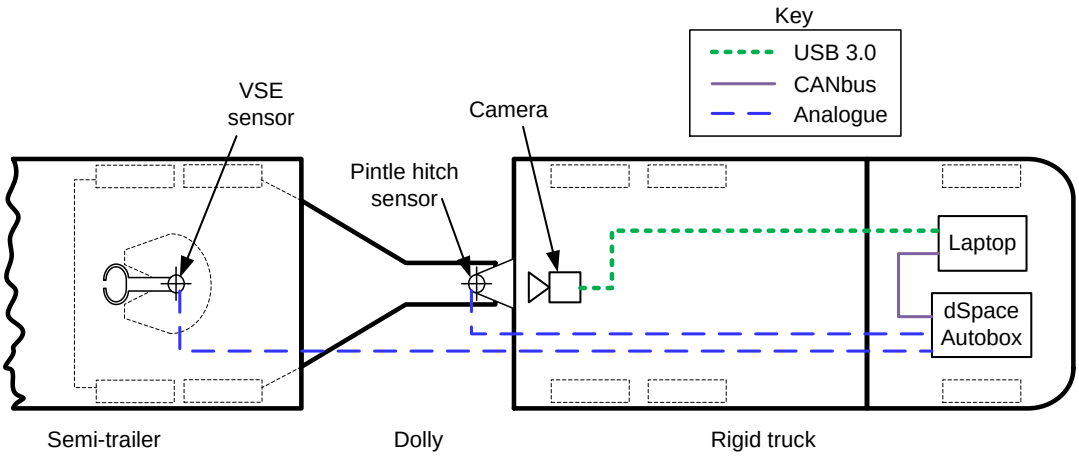


Fig. 5.3 Nordic combination instrumentation

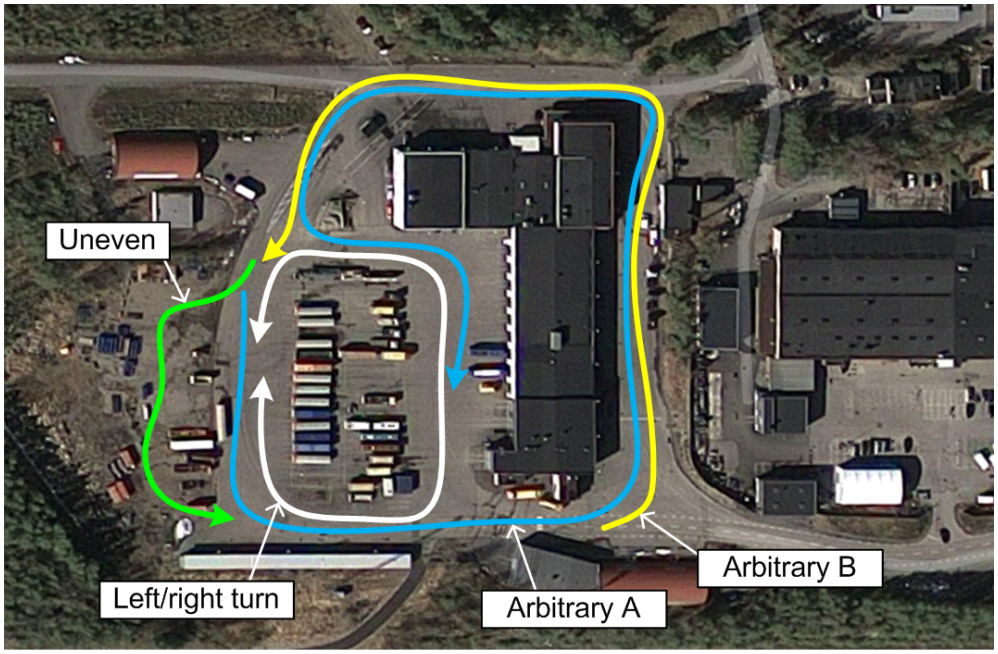


Fig. 5.4 Testing location at Hällered proving ground, Sweden, showing test manoeuvres

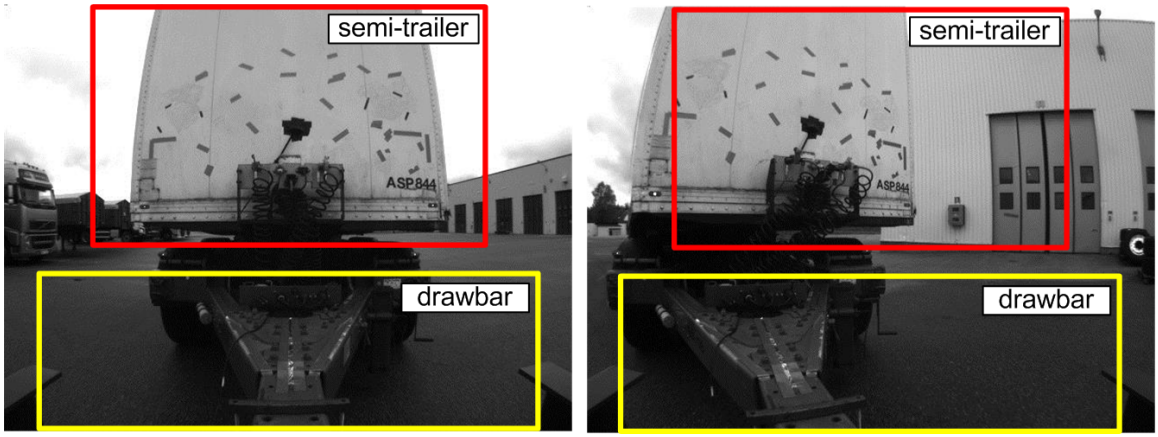


Fig. 5.5 Image regions of interest used for processing (zero and non-zero articulation angles shown)

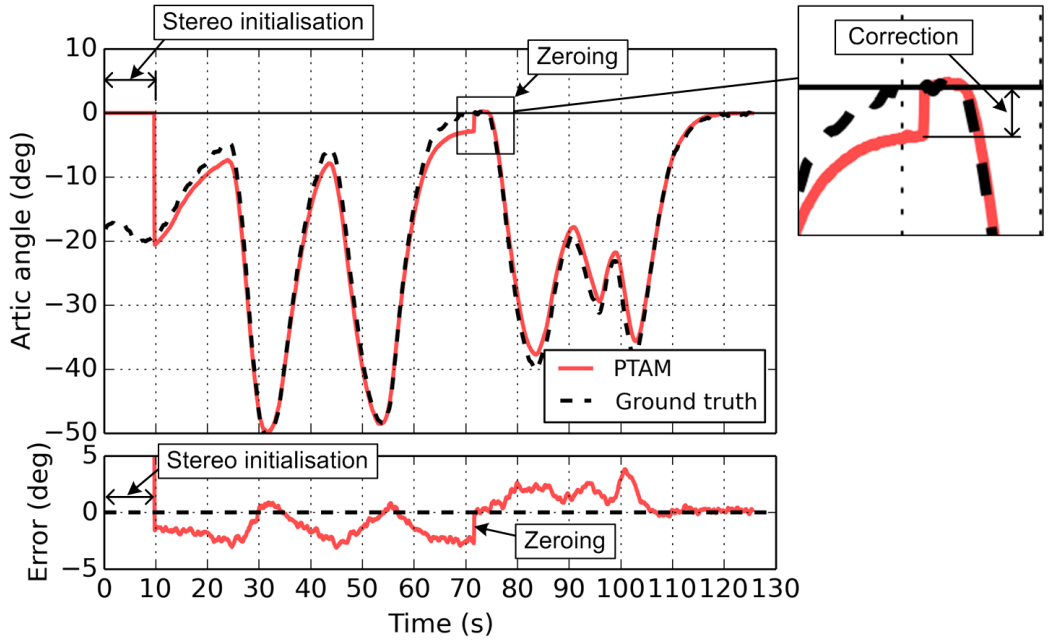


Fig. 5.6 Stereo initialisation and zeroing (semi-trailer, right turn, run 1)

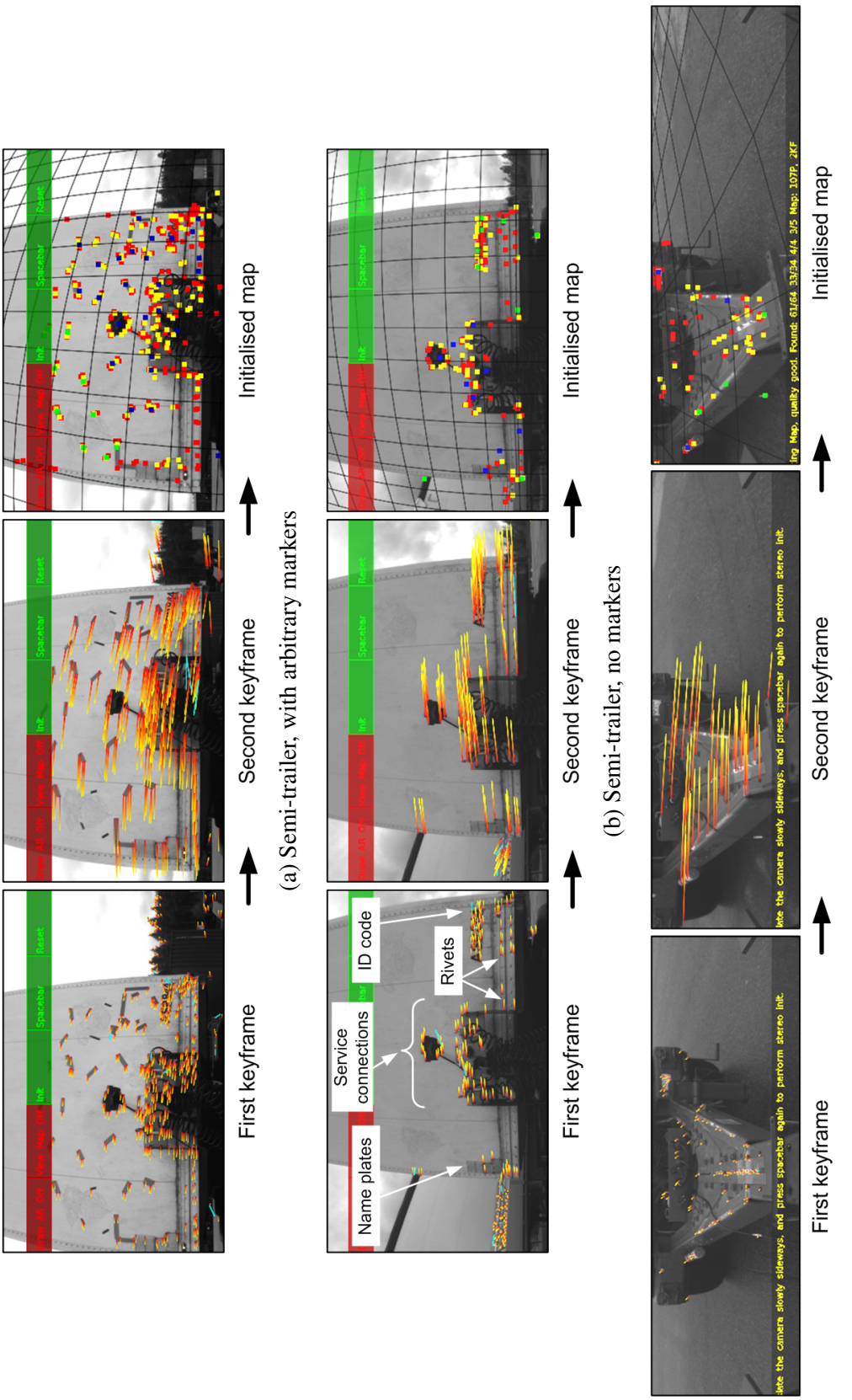
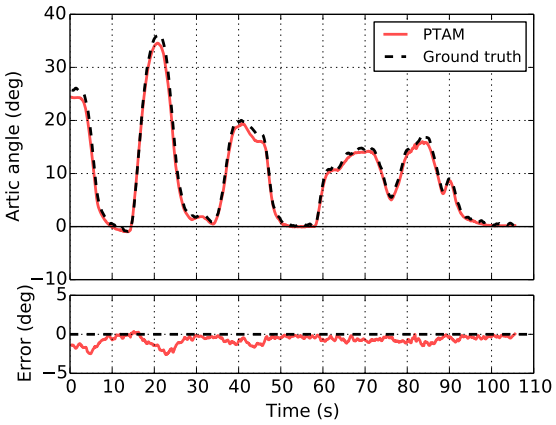
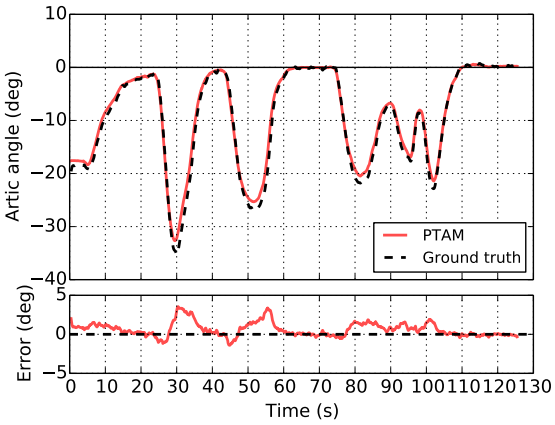


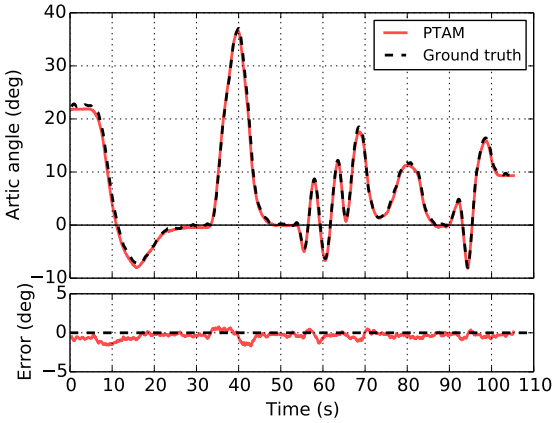
Fig. 5.7 Stereo initialisation: keyframes and initialised maps (feature tracks shown in yellow and red lines)



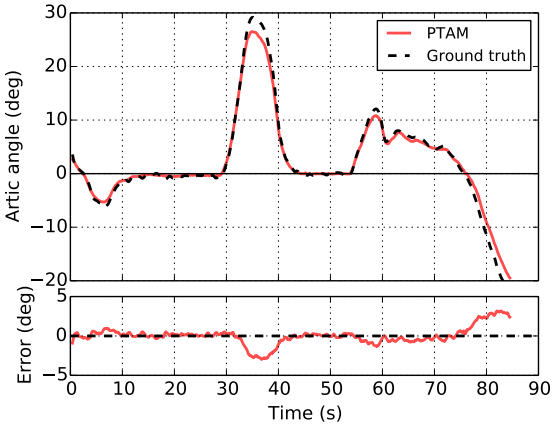
(a) Left turn



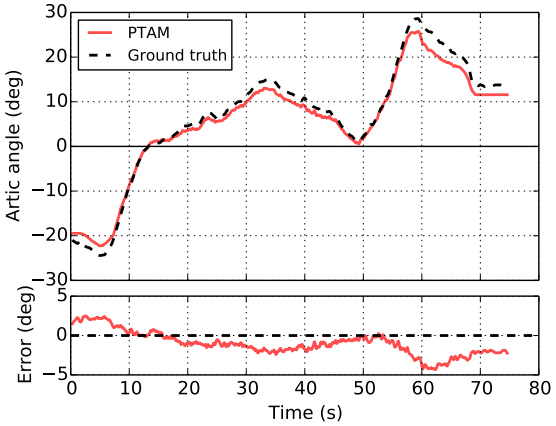
(b) Right turn



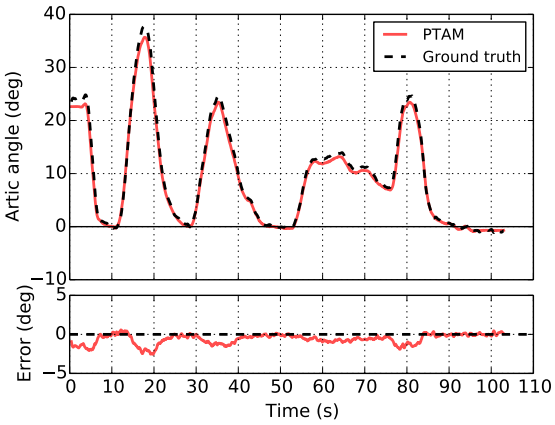
(c) Arbitrary A



(d) Arbitrary B

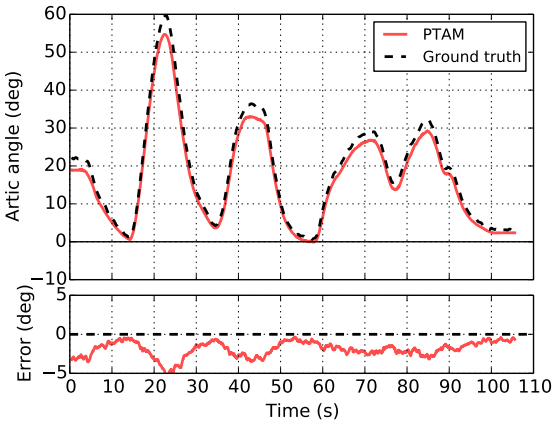


(e) Uneven

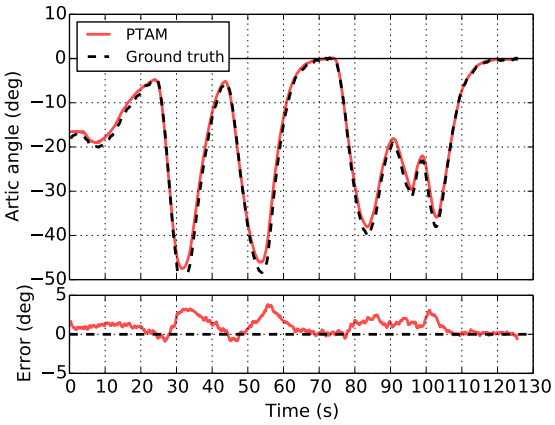


(f) No markers

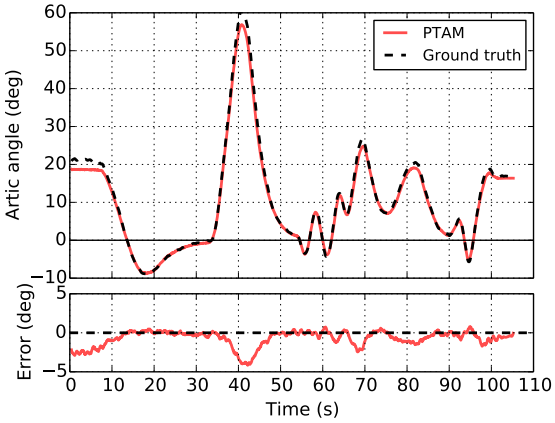
Fig. 5.8 Time histories, drawbar



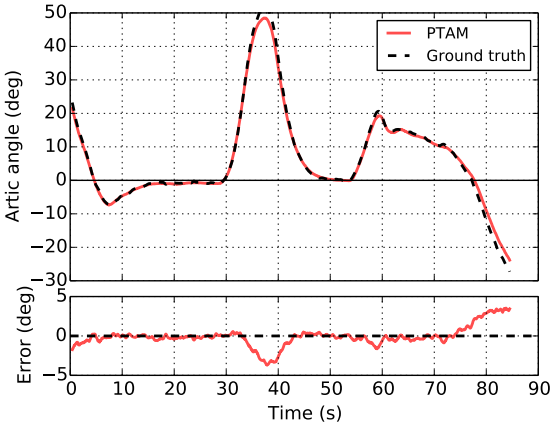
(a) Left turn



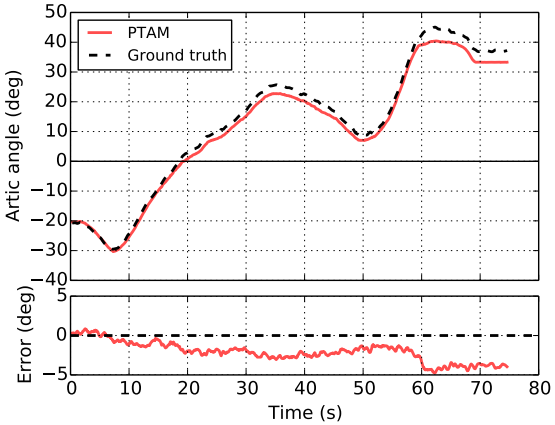
(b) Right turn



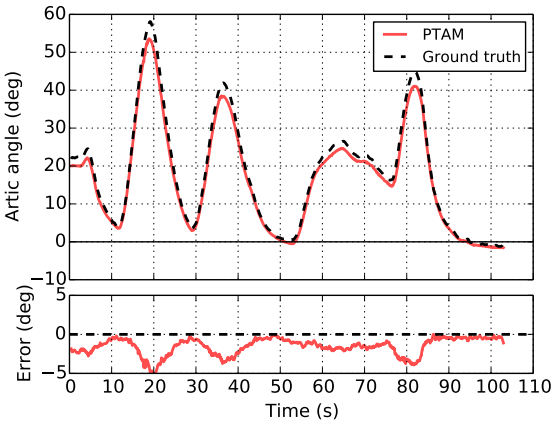
(c) Arbitrary A



(d) Arbitrary B

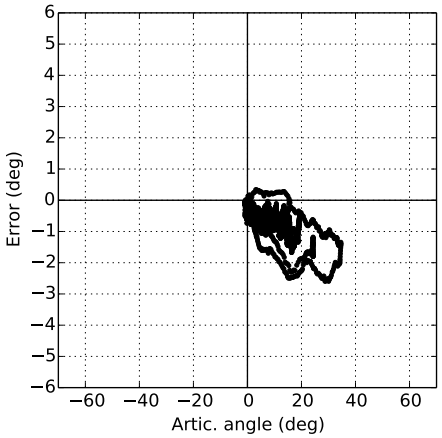


(e) Uneven

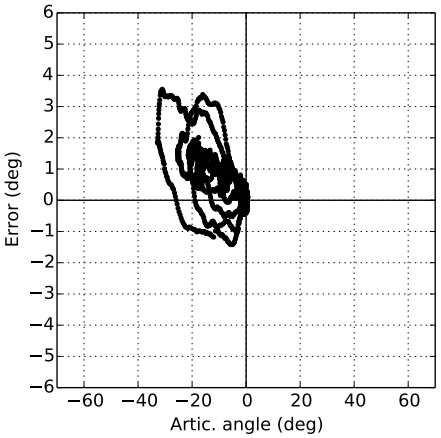


(f) No markers

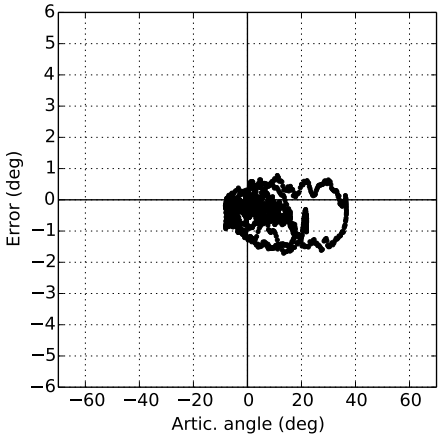
Fig. 5.9 Time histories, semi-trailer



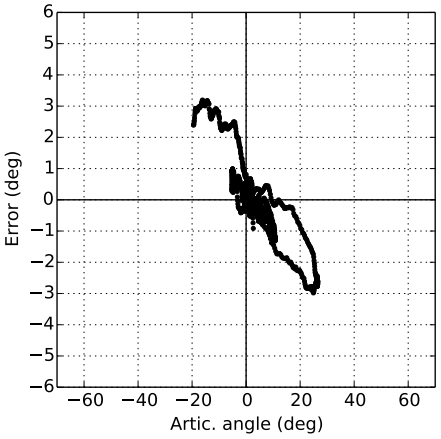
(a) Left turn



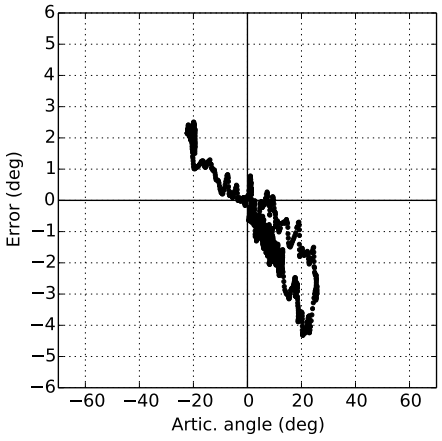
(b) Right turn



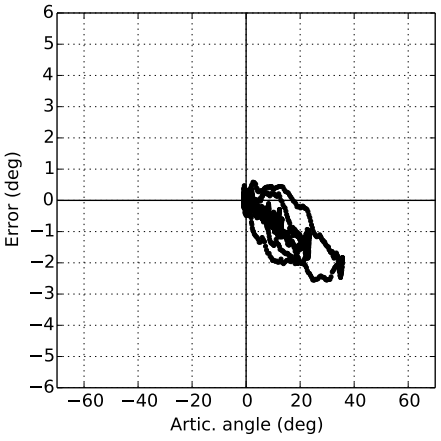
(c) Arbitrary A



(d) Arbitrary B

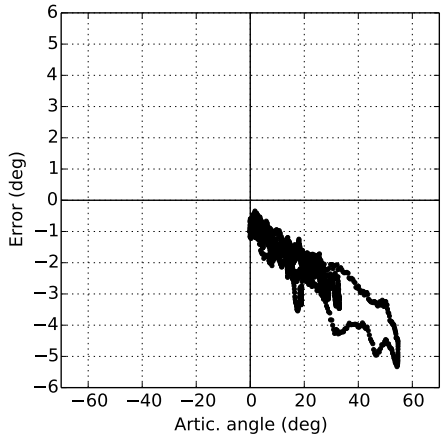


(e) Uneven

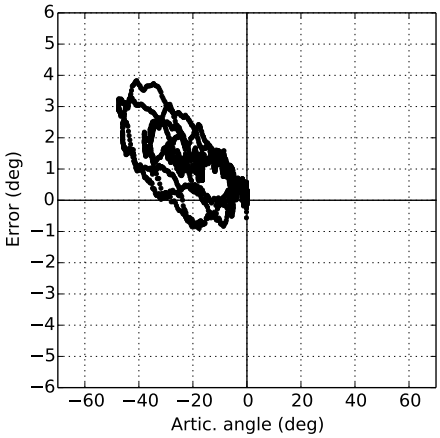


(f) No markers

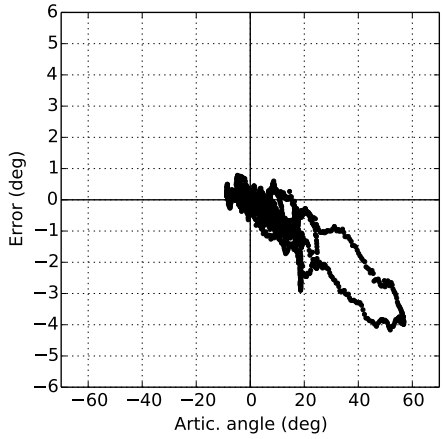
Fig. 5.10 Errors vs. articulation angle, drawbar



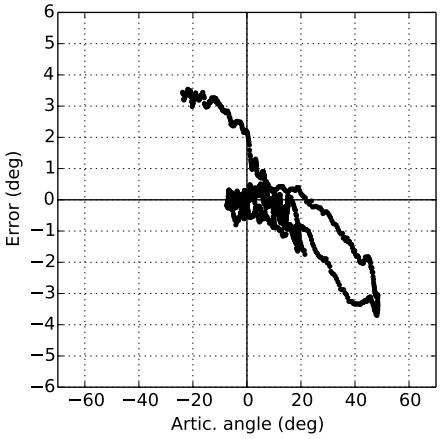
(a) Left turn



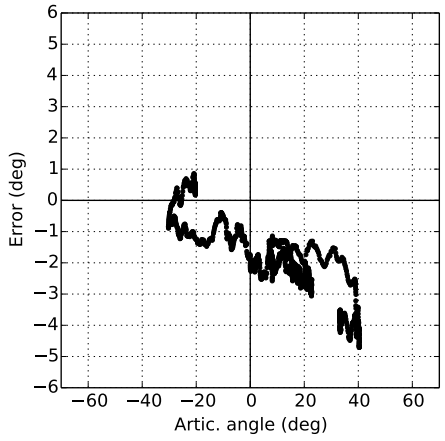
(b) Right turn



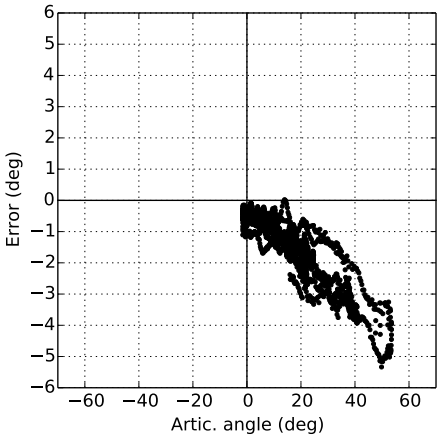
(c) Arbitrary A



(d) Arbitrary B

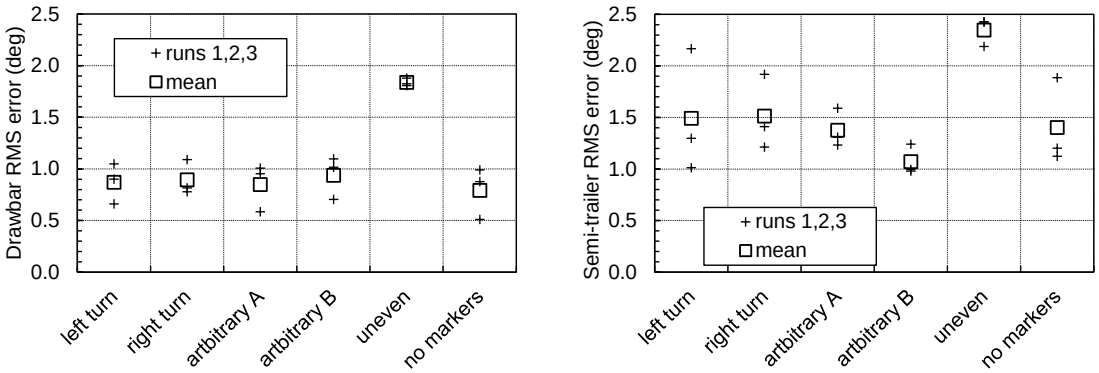


(e) Uneven



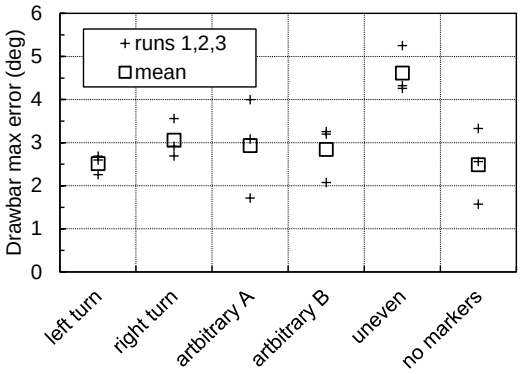
(f) No markers

Fig. 5.11 Errors vs. articulation angle, semi-trailer

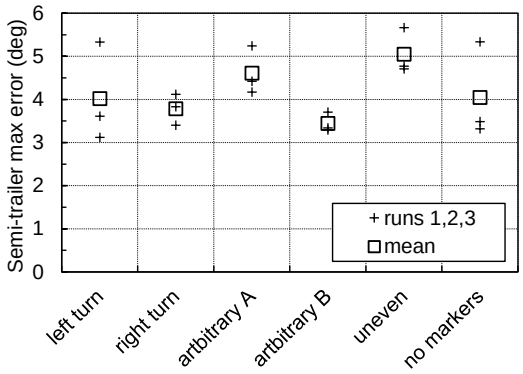


(a) RMS errors, drawbar

(b) RMS errors, semi-trailer



(c) Max. errors, drawbar



(d) Max. errors, semi-trailer

Fig. 5.12 Overview of maximum and RMS errors for all runs

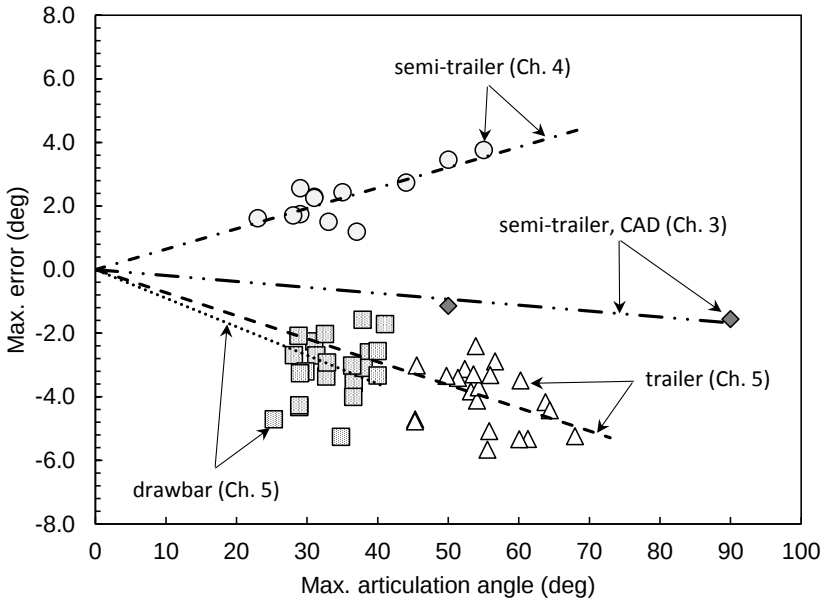


Fig. 5.13 Maximum errors vs. max. articulation angle (results from Chapters 3, 4 and 5)

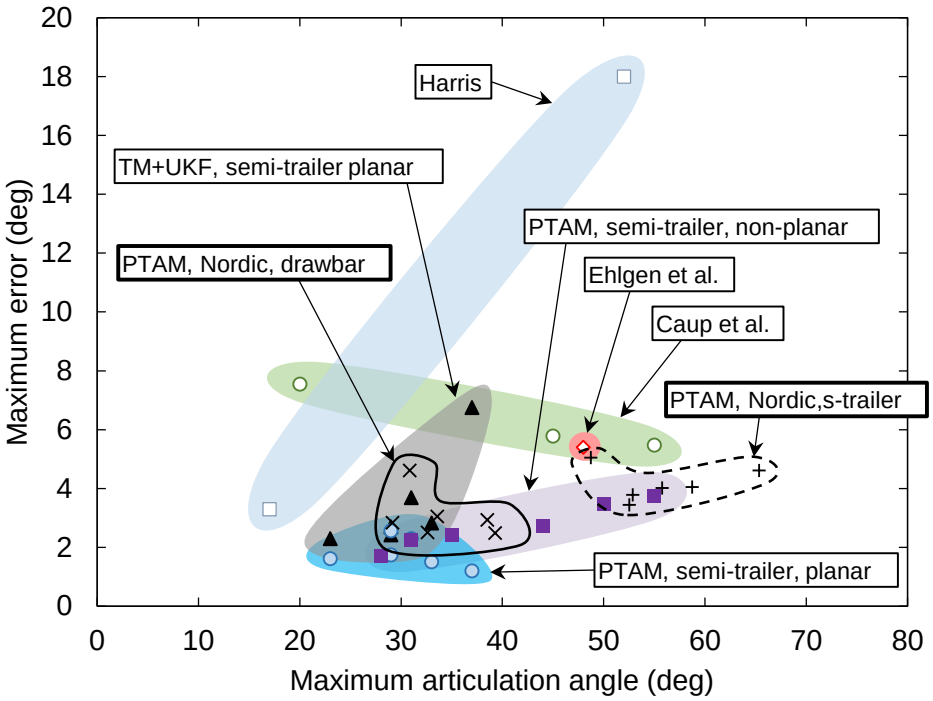


Fig. 5.14 Performance benchmark against published vehicle test data

Chapter 6

Development of a trailer off-tracking sensor

In this chapter, a trailer off-tracking measurement concept is developed, which is able to provide a direct trailer off-tracking measurement for Cheng's path-following controller [32]. The new concept addresses previous shortcomings of the controller in that it is independent of tractor-based measurements and parameters, and makes no assumptions about wheel slip.

The concept also addresses the shortcomings of Miao's ground-watching navigation system [3], in that it does not rely on a consistent ground surface from which to obtain image data. This enables the system to operate under any road conditions, including wet and muddy conditions off-highway.

6.1 Off-tracking estimation with visual odometry

A new method of trailer off-tracking measurement is proposed in this work which addresses certain limitations of Cheng's path-following controller [32] and Miao's ground-watching navigation system [3], which aims to achieve measurement errors less than 0.10 m (RMS) and 0.15 m (maximum) (Section 1.6). The method utilises a stereo camera pair mounted to the roof of a trailer (either along the side or rear edge) which captures stereo images of the surrounding landscape. These images are processed using the VISO2-S visual odometry algorithm [133] to determine the motion of the cameras (and hence trailer) relative to the surroundings. Side-facing cameras would give the system flexibility for application to multi-trailer combinations.

Trailer motion data are then manipulated to estimate trailer off-tracking by finding the relative trajectories of the fifth wheel and trailer follow point. The adopted visual odometry

algorithm is robust to outliers and to regions of moving scenery. The combined VISO2-S and off-tracking calculation algorithm is denoted 'VISO-OT'.

6.1.1 Visual odometry

PTAM was initially considered for the off-tracking estimation task, given the good performance exhibited in the articulation angle estimation task. However, PTAM relies on computationally expensive bundle adjustment for its accurate pose estimation which is best suited for applications where the map is not steadily growing. From Klein and Murray's paper on PTAM [123]:

> One way or the other, it becomes an increasingly expensive computation as map size increases: For example, tens of seconds are required for a map with more than 150 keyframes to converge. This is acceptable if the camera is not exploring (i.e. the tracking system can work with the existing map) but becomes quickly limiting during exploration, when many new keyframes and map features are initialised (and should be bundle adjusted) in quick succession.

It was hence necessary to explore alternative options. Visual odometry was selected as a suitable motion sensing solution for this application.

Visual odometry is the estimation of the pose and motion of a camera through a 3-dimensional scene. Advances in visual odometry algorithms have resulted in its widespread use in the areas of autonomous road vehicles and mobile robotics. Compared to other odometry systems such as wheel speed sensors and GPS, visual odometry offers high precision, low-cost hardware, and independence from traction conditions.

Although vehicle-based visual odometry is commonplace in autonomous vehicles (see for example [133, 186]), little work has been done with heavy vehicles, with the exception of the related work of Miao [3] and Harris [73].

Numerous visual odometry algorithms exist with varying degrees of accuracy and robustness, and which utilise various combinations of cameras and other sensors such as Lidar. The VISO2-S visual odometry algorithm of Geiger *et al.* [133] was adopted for this work. The algorithm requires sequences of stereo image pairs from a stereo camera, and assumes the stereo cameras have been calibrated and the calibration information is available. The 'stereo baseline' (the lateral separation between left and right cameras) must also be known.

Although some other algorithms offer superior measurement accuracy [131], VISO2-S does not require additional sensors and has a robust method of minimising the effect of moving scenery which is important for transport applications. Furthermore, the open-source

C++ source code is available online[1] which is attractive for this proof-of-concept work. In future development, more recent and more accurate algorithms could be explored.

Details of the VISO2-S algorithm may be found in [133] and can be summarised into the following steps:

1. A stereo image pair is obtained and corner-like features are detected in each image.
2. 'Circular' feature matching is performed, comparing features between left and right images (normal stereo matching) as well as between current and previous image pairs. Features are accepted if matching succeeds through the entire 'circular' loop of four images.
3. A 'bucketing' process [187] divides the images into a rectangular grid, and each 'bucket' may only store a maximum number of features. This ensures a good distribution of features around the image, minimising the effects of bias and of moving objects.
4. Ego-motion is estimated by minimising reprojection errors through Gauss-Newton optimisation with respect to R and **T**, the rotation matrix and translation vector respectively.
5. The ego-motion estimation incorporates a RANSAC strategy to remove outliers.
6. A constant acceleration Kalman Filter is used to minimise noise.

In performance tests on the KITTI dataset [188], the VISO2-S algorithm was shown to yield 2.44% translation error and 0.0114 deg/m rotation errors on average with a 0.5 m stereo baseline. The algorithm runs at 20 fps on a single processing core using the recommended parameter settings.

Using a representative trailer length of 14 m (from fifth wheel to follow-point), translation drift of 2.44% would result in approximately $0.0224 \times 14\ \text{m} = 0.3136\ \text{m}$ of lateral off-tracking between the fifth wheel and the rear of the trailer. It is expected that the accuracy of VISO2-S can be improved nearer to the target maximum error of 0.15 m by using a larger stereo baseline. This is discussed in more detail later.

A mono camera variation of the algorithm exists but is less accurate, less computationally efficient, and relies on scene assumptions. Mono visual odometry is also more prone to drift [82], though some drift may still exist with stereo cameras due to inaccuracies in camera calibration.

[1] http://www.cvlibs.net/software/libviso/, last accessed 20 April 2017

The use of a single camera is attractive in size- and mass-limited applications (such as drones), but is of little consequence here. Consequently, stereo odometry was adopted for it benefits in terms of accuracy, drift, and independence of any world assumptions. Drawbacks include a small increase in hardware cost, and additional calibration steps.

6.1.2 Off-tracking calculation

Given visual odometry data from a stereo camera pair fixed to the trailer, the task remains to calculate off-tracking between the rear of the trailer (the follow-point) and the trajectory of the fifth wheel. Visual odometry data are in the form of a rotation matrix, R, and translation vector, **T**, at each frame. R and **T** are relative to the prior vehicle location, and in that vehicle's reference frame.

Figure 6.1a illustrates the step-by-step (or frame-by-frame) yaw-plane motion of a tractor semi-trailer combination at frames i, $i-1$, $i-2$, $i-3$, *etc.* The camera origin has been arbitrarily assumed to be at some distance a behind the fifth wheel and b ahead of the follow-point.

The raw visual odometry data are shown in the form of Δx, Δy and $\Delta\psi$: incremental translation and rotation at each frame. Only rotations and translations in the yaw plane were considered, assuming pitch and roll motions to be negligible relative to motion in the yaw plane.

By making use of a data buffer or shift register, these data may be used to calculate off-tracking. First, depending on whether the cameras are rear-facing, side-facing, or at some intermediate angle between, translational data from the cameras must be rotated and aligned with the vehicle co-ordinate frame.

To do this the camera-to-vehicle rotation parameter is defined as ψ_{c2v}. For perfectly rear-facing cameras, $\psi_{c2v} = 180^\circ$, and for cameras pointing to the right, $\psi_{c2v} = 90^\circ$. The co-ordinates can be rotated according to:

$$\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}_i = \begin{bmatrix}\cos(\psi_{c2v}) & \sin(\psi_{c2v})\\ -\sin(\psi_{c2v}) & \cos(\psi_{c2v})\end{bmatrix}\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}_{cam,i} \tag{6.1}$$

The motion data from the cameras can then be transformed into fifth wheel motion (denoted by the subscript F for 'front') as follows:

$$\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}_{F,i} = \begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}_i + a\cdot\begin{bmatrix}\sin(\Delta\psi_i)\\ \cos(\Delta\psi_i)-1\end{bmatrix} \tag{6.2}$$

where a is the distance from the hitch to the reference point of the cameras. This is illustrated in Figure 6.1b. These data are stored in a buffer which initially grows with each new frame.

To find the fifth wheel trajectory *relative to the current vehicle reference frame*, previous data in the buffer must be rotated at each new frame. This can be done by first summing $\Delta\psi$ from the current frame to each previous frame as follows:

$$\psi_{i-k}^{(i)} = -\sum_{j=1}^{k} \Delta\psi_{i-j}^{(i-j)}, \quad k = 1,2,3,\ldots \tag{6.3}$$

The counter, k, denotes the entry in the buffer, increasing from the current frame (k=1) backwards. The superscript (i) denotes the co-ordinate frame (*i.e.* that of vehicle i).

For buffer entry k, and hence frame $i-k$, a rotation transformation from frame $(i-k)$ to frame (i) can then be applied to obtain Δx and Δy in the current reference frame:

$$\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}_{F,i-k}^{(i)} = -\begin{bmatrix}\cos(\psi_{i-k}^{(i)}) & \sin(\psi_{i-k}^{(i)})\\ -\sin(\psi_{i-k}^{(i)}) & \cos(\psi_{i-k}^{(i)})\end{bmatrix}\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}_{F,i-k}^{(i-k)} \tag{6.4}$$

which is illustrated in Figure 6.1c.

The x and y locations of each data point in the current vehicle reference frame (centred on the fifth wheel) may then be found by simple summation of Δx and Δy data:

$$\begin{bmatrix}x\\ y\end{bmatrix}_{F,i-k}^{(i)} = \sum_{j=0}^{k-1}\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}_{F,i-k}^{(i)} \tag{6.5}$$

This result is illustrated in Figure 6.1d.

At a given frame i, (x,y) data of fifth wheel path history points allow us to find the off-tracking at the rear of the trailer. Assuming the dimension $a+b$ to be known, the most recent path history point for which

$$y_{F,i-k}^{(i)} \leq -(a+b) \tag{6.6}$$

denotes the point beyond which the fifth wheel trajectory is past the rear of the trailer.

The value of k at which this condition is met is denoted k_b, or the buffer length. The buffer only needs to retain k_b data points, with older data being discarded and the most recent data added at each new frame in the data buffer.

The off-tracking, e_{tr}, may be calculated by linear interpolation as follows:

$$e_{tr} = x^{(i)}_{F,i-(k_b-1)} + \left(x^{(i)}_{F,i-k_b} - x^{(i)}_{F,i-(k_b-1)}\right)\left(\frac{y^{(i)}_{F,i-(k_b-1)} + (a+b)}{y^{(i)}_{F,i-(k_b-1)} - y^{(i)}_{F,i-k_b}}\right) \tag{6.7}$$

If at least one trailer length has passed, the above calculation of off-tracking will be possible. If not, the buffer will not contain sufficient data for off-tracking to be calculated. In this case the next frame is obtained, and the buffer simply grows in size until sufficient data are available and off-tacking can start to be calculated. Thereafter, the buffer size will vary with the speed and path of the vehicle, only keeping data old enough to calculate off-tracking.

When a vehicle stops, it is possible to store the contents of the buffer. So in theory, the calculation of a semi-filled buffer should only occur the first time a new vehicle combination is launched.

Integration drift is inherent in visual odometry systems as a result of summing incremental motion data in the above fashion. In typical automotive and mobile robotics applications, these data are integrated indefinitely with time to obtain global position estimates, so the integration errors can grow unbounded. However, in this application integration is performed only from the fifth wheel to the rear of the trailer and so integration drift is bounded by the length of the trailer and will not grow indefinitely with time. Further, the effects of any outliers in the visual odometry data will be removed from the buffer after one trailer length has passed.

6.1.3 Wide-baseline stereo

In Section 6.1.1, it was theorised that the accuracy of VISO2-S could be improved by using a larger stereo baseline in an attempt to approach the desired error targets of 0.10 m (RMS) and 0.15 m (maximum).

Increasing the baseline of a stereo camera pair can improve depth accuracy and hence odometry accuracy [189]. However, the extent to which this can be utilised is limited due to the increasing difference in perspective between the two cameras. This will reduce the quality of feature matching and hence reduce the total number of successful matches between stereo image pairs. For this reason standard stereo vision algorithms are sometimes adapted for wide-baseline applications primarily by using more robust (but computationally expensive) feature similarity metrics (see for example [190, 191, 189]).

There is also a practical limitation for vehicles and mobile robots, where the camera baseline cannot practically exceed the dimensions of the vehicle or robot, or even some

proportion of it. For passenger cars, the stereo baseline is also limited by more convex shapes of the bodywork, and by aesthetics.

For heavy goods vehicle applications, particularly on trailers, these constraints are less problematic. Shapes tend to be rectangular and perpendicular, and dimensions are significantly larger.

The most limiting dimensional constraint for rear-facing cameras would be that of maximum vehicle width, which in the UK is legislated to 2.5 m. Larger baselines are possible parallel to the sides of the vehicle for side-facing cameras. However, other practical limitations become important in these cases, concerning ease of stereo camera calibration and the rigidity of the camera mounting.

Wide baselines were considered as part of this investigation. In this case the scenery was assumed to be sufficiently distant that no alterations to the stereo matching algorithm were necessary.

6.2 Simulation overview

As a proof-of-concept, and to assess the theoretical accuracy of the VISO-OT system, simulations were carried out in Autodesk Inventor [167] in which visually representative road and roadside environments were constructed. A stereo camera pair was made to travel through the scene while rendering images from each camera at a fixed frame rate. VISO-OT was then used to process the image sequences and estimate off-tracking.

The 'virtual environment' is shown in Figure 6.2a. It included a gravel road, grassy roadsides, trees, shrubs, fences, and distant clouds. Road width was set to 5 m (within the UK rural road design guidelines), and the size of objects and textures were chosen to be representative. Soft ambient lighting and shadows were incorporated and all scenery was stationary. The scenery repeated every 100 m along the road.

A stereo camera pair of set baseline was made to follow a straight path along the road at a given slip angle, representative of a trailer moving with constant off-tracking due to a cambered road surface for example. This is illustrated in Figure 6.2b.

Stereo image pairs were rendered at each frame using Autodesk Inventor Studio, and sequences of image pairs were obtained at 10 fps as the trailer travelled 100 m along the path at a constant speed of 5 m/s.

Images were captured at a resolution of 1344×391 with a 100° field of view. The cameras were located 3 m above ground level with zero tilt or roll angles relative to the ground. As in Chapter 3, the intrinsic cameras parameters were determined without the need

for camera calibration, using only knowledge of the image size and field of view (Equations 3.13). In this case: $f = 563.9$ pixels, $(u_0, v_0) = (671.5, 195.5)$ pixels.

Each generated image sequence was post-processed using VISO-OT. Off-tracking measurements were made, and metrics such as RMS error, mean error, mean number of feature matches, and mean number of inlying feature matches were recorded for each 100 m run.

Batch processing of the simulations was achieved using an AutoIt script (open-source script language for automation of tasks in Windows [192]). The script was used to:

1. update run parameters in an Excel spreadsheet, which were dynamically linked to Autodesk Inventor,

2. perform the image rendering task in Autodesk Inventor and save image data to file,

3. execute a Python script to greyscale the images, and

4. execute the VISO-OT algorithm on the image sequence and write results to file.

For step 4, a MATLAB-based variant of the visual odometry algorithm was used for simplicity. The algorithm is fundamentally identical to the C++ implementation. All CAD and visual odometry processing was performed on the same 3.2 GHz desktop computer used previously (see Appendix A for specifications).

The default input parameters for VISO2-S were used, including 50 RANSAC iterations per optimisation, outlier flow and disparity thresholds of 5 pixels, a bucket size of 50×50 pixels and a maximum of 2 features per bucket.

For off-tracking calculations, the trailer length from fifth wheel to follow-point ($a + b$ in Figure 6.1) was taken to be 14 m. This is approximately representative of a UK longer semi-trailer (LST). Given the trailer side-slip angle and knowing the trailer length, the 'ground-truth' off-tracking was known accurately in each case for comparison.

The simulation environment was devised so that many of the described simulation and camera parameters were variable. These included trailer slip angle, camera location (*e.g.* rear or side-mounted), stereo baseline, image resolution, field of view, run length and frame rate. This allowed for sensitivity to variations from the 'reference' settings to be investigated.

Four broad categories of simulations were conducted:

1. a sensitivity study, to confirm that simulation results were not sensitive to limitations of the simulation environment such as run length and the size of the virtual landscape;

2. a study of the effect of camera parameters, to confirm the choice of camera resolution and field of view;

3. a study of the impact of varying the primary design parameters, such as camera baseline and camera configuration;

4. a study of the performance of the system under a range of operating conditions such as trailer slip, scenery density, and vehicle speed.

A reference parameter set was chosen from which variations were investigated. The reference parameters were as follows: 14° trailer slip, 100 m longitudinal travel, 2.5 m lateral offset from the road edge, a starting location of 50 m from the edge of the simulation environment, 10 renders per image, 100° field of view, an image resolution of 1344×391, and a frame rate of 10 fps. The reference parameters (50 m, 100 m, 10 renders) were chosen to give a good balance of perceived realism, while ensuring a reasonable simulation time.

Note that unlike the articulation angle simulation model (Section 3.1), the current simulation model is purely visual. Steady-state motion in a straight line is assumed, with constant side slip and speed, removing the need for a dynamic vehicle model.

6.3 Results

Example camera views from the stereo camera pair are shown in Figure 6.3 for both rear and side facing camera cases. Locations of matched features are shown in each image. The example case shown is for a trailer slip of 10° and a baseline of 2.5 m (to maximise the differences in perspective). Detailed results of the four simulation studies are presented in the following sections.

6.3.1 Simulation design

Although it is clear that the simulation is an idealisation, it was important to assess that it was not biased on account of limitations of the simulation environment, such as the distance travelled through the repeating scenery. As such, a few simulations were conducted to assess the sensitivity of a chosen reference case to some of the main simulation parameters.

Firstly, sensitivity to the following simulation parameters were assessed: longitudinal offset (distance from the rear edge of the scene at the start of the run), travel distance, and image renders (the number of computational passes to create realistic lighting and textures in the scene, with increasing realism). This ensured that these did not have a systematic effect on the results.

Results for these simulations are shown in Figures 6.4, 6.5 and 6.6. Each of these figures shows the effect of the selected parameter on RMS error, mean error, mean number of feature

matches, and percentage of matches that were inliers. All metrics were calculated over the total travel distance for a run. Results are shown for both rear and side camera configurations.

In Figure 6.4 no systematic effects are observed due to variation of longitudinal offset on RMS errors (which averages between 2 and 3 cm) or bias, nor in the feature matching statistics. This suggests that the chosen reference value is suitable.

When considering the length of the simulation run it was expected that as run length tended toward infinity, errors would asymptotically converge to a minimum due the increasing number of data points and the nature of the RMS and mean error calculations. Over a short range of variations, some random variations in these error metrics would be expected. Conducting infinitely long simulations would be impractical. The range considered here of 100 to 250 m was chosen to be representative of a typical HGV manoeuvre.

The results for run length variations are shown in Figure 6.5. Here we also observe no systematic effect of changing the run length from 100 m to 250 m, and so 100 m was deemed to be suitable for these simulations to minimise simulation time.

Increasing the number of image renders used by Autodesk Studio increases the realism of the textures and shadows of the images, but at significant computational expense. Theoretically, infinite image renders would give the most realistic rendering of textures and shadows, minimising any misleading artefacts as a result of insufficient rendering.

The sensitivity to number image renders, shown in Figure 6.6, shows that 10 renders seems to be a good compromise. Using less than 10 renders gives noticeable increases in RMS and mean errors, and reductions in features matched, while using more than 10 has only a small effect on these metrics but significantly increases rendering time (not shown).

Sensitivity to the lateral offset of the vehicle from the road edge was also considered, partly to validate the simulation design but also to give insight into the effect of the average distance to features. These results are shown in Figure 6.7. Although some noise is present in the side camera results, there is no evident effect on errors. There is only a small increase in features with increasing lateral offset, expected due to the increasing similarity of features. Overall the selected offset is suitable and unbiased.

From these results, it can be concluded that the selected reference parameters for longitudinal offset, run length, lateral offset and image renders were suitable, while keeping simulation times practical. On average, it took 23 minutes render a full image sequence per run on a desktop computer with a 6-core[2] 3.2 GHz CPU and 16 GB of RAM (see Appendix A for details).

[2]The 2016 version of Autodesk Inventor supports multi-threading for some operations including rendering

6.3.2 Optical parameters

Sensitivity to camera design parameters was assessed to validate the chosen values and assess if there was room for improvement. The reference camera parameters of 100° field of view and a resolution of 1344×391 were based largely on previous work (for example [188]), in which they were found to offer a good balance between maximising the view and pixel resolution in the yaw plane, whilst reducing the computational demand.

Results for resolution sensitivity are shown in Figure 6.8. Note that the aspect ratio of the reference 1344×391 was retained when altering the resolution. Also, the bucketing parameter of 2 features per 50×50 bin was retained, meaning that increased resolution was expected to give a direct increase in the number of matched features. The increase in features with resolution is clear, but the percentage of inliers stays approximately constant, indicating that the 'quality' of the matches does not necessarily improve with better resolution.

A reduction in errors is clear from low resolutions of around 500 pixels (width) towards the reference value of 1344, but little improvement is in evidence thereafter. These findings suggest that this resolution remains a good compromise, because further increases would negatively impact frame rate, but with little benefit in terms of accuracy.

Sensitivity to the field of view is shown in Figure 6.9. Here we observe that the the value of 100° is almost optimal, giving close to the lowest errors and maximum number of feature matches. In a practical sense, a 100° field of view gives a reasonably wide view, but without approaching the severe distortions of wider angle lenses (such as fish-eye lenses). More severe lens distortion would require more computational for undistorting the images before processing. So this finding is favourable, confirming the choice of the reference parameter.

6.3.3 Design parameters

One of the primary system design variables is the stereo baseline. It was hypothesised earlier that a wider baseline might improve the accuracy of the visual odometry measurements relative to the published accuracy in [188]. A range of baselines were considered against the reference case, from 0.5 to 5.0 m in 0.5 m increments. A practical limit of 2.5 m would exist for rear-facing cameras given the width of the trailer, but larger baselines were explored for potential use on the side of the trailer.

The results are shown in Figure 6.10. A clear reduction in feature matching is evident with increasing baseline, as expected due to the increasing difference in perspective of the two cameras. However, little effect on errors is evident until a baseline of approximately 2.5 m. Until this point, RMS errors are in the region of 1–4 cm, and mean errors in the order of 1 cm.

The reduction in the number of feature matches and increase in errors are more pronounced for the side-facing cameras. This is expected due to the smaller average feature depth which is less suited to wider baselines. The point at which errors seem to grow (at a baseline of 2.5 m) appears to correlate with the number of feature matches dropping below 250 (shown in the figure). This may be indicative of a useful practical limit when field-testing the system.

At reasonable baselines (below 2.5 m), the results for rear and side cameras are comparable. However, for higher baselines, the side cameras experience a significant increase in RMS errors compared with the rear-facing cameras. The average feature depth for the side cameras is lower, and an increased baseline gives better feature matching for more distant features. So an increased baseline will have more effect on the side-facing cameras.

Additional variations in camera orientation were investigated for rear cameras at a 500 mm stereo baseline, beginning with rear cameras ($\psi_{c2v} = 180°$), and assessing incremental rotations towards $\psi_{c2v} = 135°$ relative to the trailer longitudinal axis. A 45° range of slip angles was assessed as well. The results are shown as contour plots in Figure 6.11. Note that mean errors are shown as absolute values for clarity.

Allowing for some random variations in results from one test to another due to noise, the results suggest that small rotations towards the side result in only minor variations in accuracy and feature matching performance. It appears that variations of up to 20° are acceptable for low slip angles (see the mean error results in particular).

However, for rotations further from rear-facing, errors are more prone to increase with increasing trailer slip, as the cameras are more perpendicular to the motion of the trailer and hence more like side cameras. For example, the results for a camera offset of 165° (15° from rear facing) are similar to the rear facing case at zero slip. However, at a slip of 15°, where there is little effect on the rear facing cameras, errors increase approximately two-fold for the 15° offset cameras.

Rear cameras therefore offer the best accuracy and consistency over the anticipated operating range of off-tracking.

6.3.4 Performance under varying conditions

In addition to the reference case of 14° trailer slip (about 3.5 m off-tracking), a wide range of other slip angles and hence magnitudes of off-tracking were investigated. Slip angles of between 0° (0 m off-tracking) and 60° (24 m off-tracking) were considered. Although off-tracking beyond approximately 4 m is not realistic, these higher slip scenarios were included for possible additional insights.

The results are shown in Figure 6.12 for a rear-mounted camera configuration. Results for three baselines—0.5 m, 1.5 m, 2.5 m—are shown for comparison.

For realistic slip angles below 16° (4 m off-tracking), the effectiveness of the feature matching is seen to be approximately insensitive to slip angles in this range, but clearly sensitive to baseline. RMS errors for the 0.5 and 1.5 m baselines are comparable, and range from 0.01 m to 0.03 m. 2.5 m baseline errors are consistently higher: between 0.02 and 0.04 m. Mean error trends are similar, with a maximum bias of 0.01 m for the 0.5 m and 1.5 m baselines, rising to up to 0.035 m for the 2.5 m baseline case.

The mean number of feature matches for the 0.5 m and 1.5 m baselines are approximately 380 and 325, well above the threshold of 250. However, for the 2.5 m baseline, the number of matches is on the limit of this threshold, sometimes dropping below it. This explains the larger RMS errors for the 2.5 m baseline, compared to those for the 0.5 m and 1.5 m cases.

These results are promising given the sub-0.15 m target accuracy. However, these results are only an indication of the limit of performance in an idealised environment, with perfectly calibrated cameras, and consistent lighting and scenery conditions.

The scenery used was selected to be as representative as possible, especially in the texture and scale of the road and grass ground surfaces. The density and distribution of 3-dimensional objects such as trees and shrubs were selected by eye to give an appropriate scene. It was not possible to determine a truly 'representative' scene given the wide variety of possibilities, but this was deemed sufficient to provide an appropriate number of visual features.

Sensitivity of the measurements to the density of the surrounding 3-dimensional scenery was assessed in order to give insights into real world performance under low feature environments. Three scenarios were considered. The reference case was deemed to have '100%' of the available scenery. A 50% case was then considered, wherein 50% of the 3-D objects were removed from the scene, at an approximately even distribution in the scene. Lastly, a 0% scene was considered in which all the 3-D scenery was removed, leaving only the ground textures. These are shown in Figure 6.13 and Figure 6.14 for the rear and side cameras.

The results for the three scenery scenarios are shown in Figure 6.15. As expected, the number of features drops with a reduction in scenery, for both rear and side cameras. The effect on errors however is distinctly different between the rear and side cameras. The rear cameras appear almost insensitive to the changes, which is a significant finding. This suggests that in typical off-highway scenery such as large crop fields or pastures, the algorithm is still able to detect and match sufficient features to maintain accuracy. This suggests that the most important feature points are ground-based, and so having a good view of the ground and road surface will be beneficial for performance.

In the case of the side cameras however, there is a two-fold increase in RMS errors when the 3-dimensional scenery is removed. Because the number of features in both rear and side cases is above 250, performance here is largely dictated by the average depth to feature (depth errors are smallest for closer features).

With side cameras the average feature depth changes more when the scenery is removed than in the rear cameras case. For the side cameras the scenery is closer to begin with and more dominant in the field of view, and so the effect of removing it is more severe, resulting in the observed rise in errors.

The existing algorithm of [133] is able to run at 20 fps, given similar image and visual odometry settings used here. A framerate of 10 fps was used for the current simulation task, and helped to reduce simulation rendering time. However, sensitivity to frame rate was explored briefly. By varying the frame rate in the simulations, and keeping all other parameters constant, the net effect is to vary the distance travelled between frames. This is hence also representative of varying speed.

Results are shown in Figure 6.16. Frame rates of between 1 and 20 fps were explored, which is equivalent to speed variations between 50 m/s and 2.5 m/s. Above 10 fps, there is no discernible effect on errors or feature matching. Below 10 fps, a gradual reduction in feature matches becomes evident, but with little effect on errors. However, for frame rates below approximately 2 fps, a significant drop in features and increase in errors is clear. Again, this point seems to correlate with the 250 feature matches threshold.

At 2 fps the separation between frames is 2.5 m, the effective ‘stereo baseline’ in the time domain. (Recall that feature matching is conducted in a circular manner, matching both left and right images, as well as current and previous frames.) This value correlates with the 2.5 m baseline limit observed in Figure 6.10. We can hence conclude that baselines and frame separations of no greater than 2.5 m should be used, to maintain sufficient feature matching performance (above 250 features). In practice, if large speed variations are anticipated, a variable framerate could be used to maintain suitable distances between frames. This may be most useful at very low speeds or when stationary to ensure a minimum distance between frames to minimise drift.

6.3.5 Bounded drift

An example time history of off-tracking error as a function of distance travelled is shown in Figure 6.17a, at 7.5° slip and 2.5 m baseline. The reference value is shown as a dashed line. No cumulative drift is apparent in these or any of the other results. Instances of temporary error drift were observed in some results however, for example in Figure 6.17b (side cameras, 5° slip, 1.5 m baseline).

Drift develops in region 'A', resulting in a constant error in region 'B'. The source of this can be seen in Figure 6.17c, in the visual odometry data in the camera Z_c-direction, where Z_c in this case is in the direction of off-tracking measurement (as is the case with the side cameras). While the data exhibit predominantly zero-mean noise, in region 'A' there is a distinct sequence of biased outliers (circled) relative to the dashed reference value.

The sum of the magnitudes of these outliers equates to a cumulative error of about 0.08 m, which is comparable to the observed off-tracking error in region 'B'. The effect disappears after approximately one trailer length has passed (14 m) as the corrupting data points are discarded from the buffer.

Although ideally no drift would be present, the temporary or 'bounded' nature of the drift here is a significant benefit over using visual odometry methods for global positioning estimates. In this case the drift was well below the maximum acceptable error of 0.15 m, but only in these idealised conditions.

Depending on the tolerances of the application, it would be desirable to try to detect or otherwise minimise drift in practical applications of the system.

6.4 Conclusions

1. A novel concept for measuring the trailer off-tracking of articulated goods vehicles was developed, using a stereo camera pair affixed to either the side or rear of the trailer or semi-trailer. The concept utilises the visual odometry algorithm of [133], combined with a buffer-based off-tracking algorithm.

2. The VISO-OT concept addresses the limitations of Cheng's path-following controller in that it is independent of vehicle slip conditions and requires no tractor-based measurements or parameters. Only the location of the cameras on the trailer needs to be known. It also improves on Miao's ground-watching system in that it does not assume a planar and unchanging road surface, making this concept better suited to off-highway applications.

3. The performance of the system was assessed in a virtual off-highway environment. Simulations provided an upper bound on the expected accuracy of the system, and illustrated some of the sensitivities of the system to both internal and external variables.

4. The selected camera resolution of 1344×391 and 100° field of view were shown to be good choices for this application.

5. Increasing the stereo baseline in an attempt to improve accuracy yielded a drop in feature matches. The effect on accuracy was small below baselines of 2.5 m (at approximately 250 features), but errors increased notably for higher baselines. Side cameras were more negatively affected than rear cameras. It is theorised that increasing the baseline will have a positive effect on system robustness in vehicle tests.

6. A rear-facing camera configuration was shown to yield the best accuracy in terms of RMS and mean errors, with decreasing performance as cameras are mounted more towards a side-facing configuration.

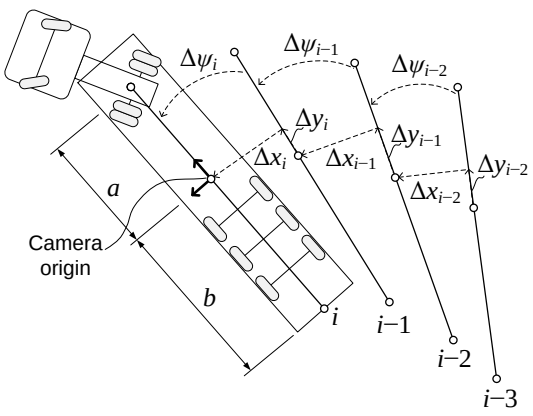
7. Under a realistic range of trailer off-tracking, in the range 0–4 m, RMS errors of between 0.01 and 0.03 m were observed for 0.5 m and 1.5 m baselines, and between 0.02 and 0.04 m for a 2.5 m baseline.

8. The system was shown to be approximately insensitive to reduced scenery density for a rear camera configuration, even in the complete absence of 3-dimensional structure (apart from texture in the ground cover). Side cameras exhibited a two-fold increase in RMS errors from a high-density scenery to none.

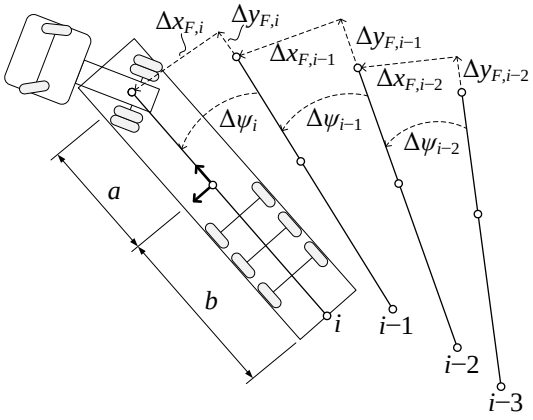
9. The algorithm was shown to be approximately insensitive to speed and frame rate, provided the distance travelled between frames was less than approximately 2.5 m (again, equating to around 250 features).

10. A small amount of drift behaviour was observed in off-tracking measurements, but this was shown to be bounded in time due to the relative nature of the off-tracking measurement and the use of a data buffer.

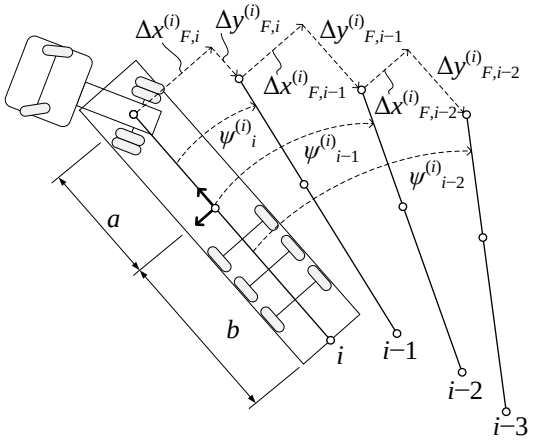
6.5 Figures



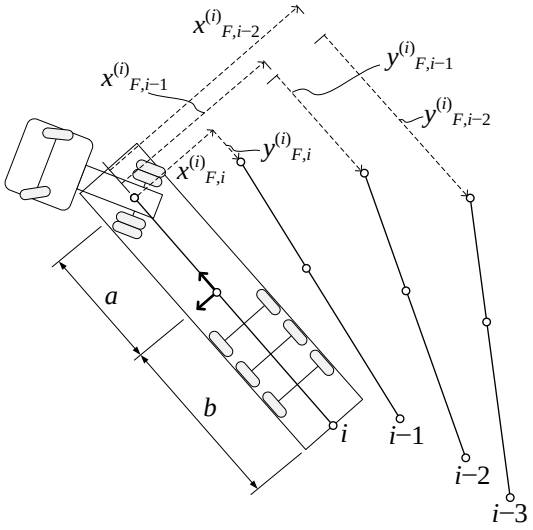
(a) Camera trajectory from visual odometry



(b) Fifth wheel trajectory

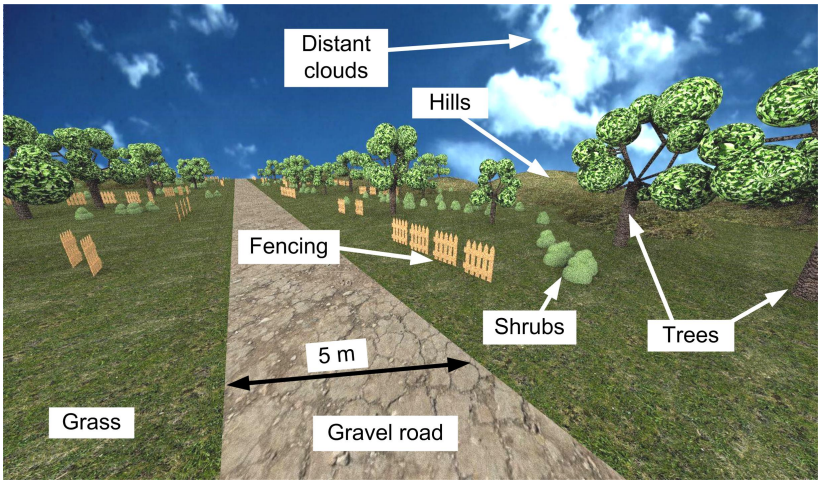


(c) Rotated trajectory in i^{th} reference frame

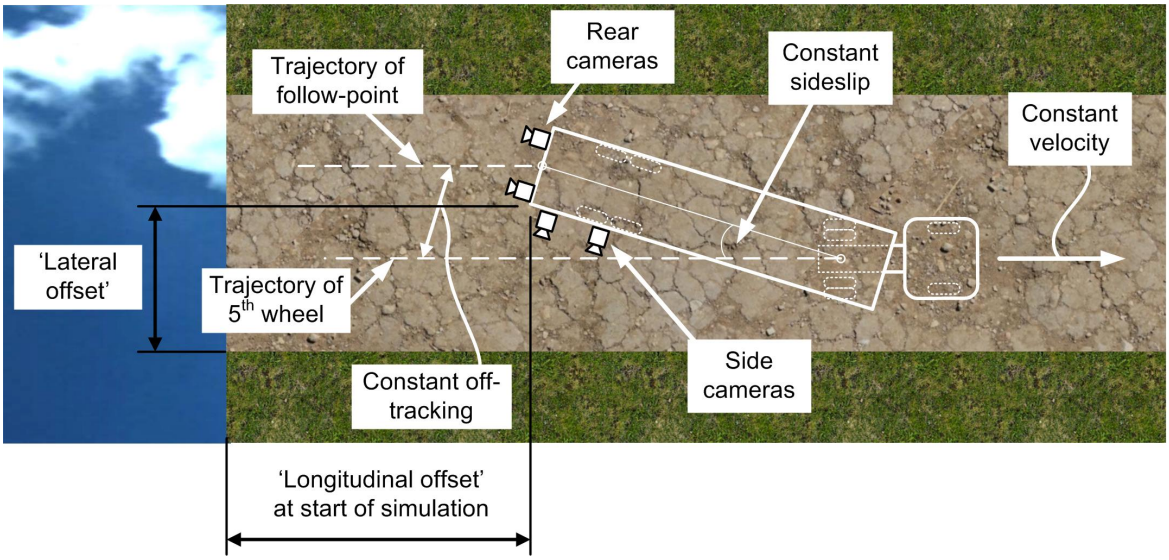


(d) x-y co-ordinates in i^{th} reference frame

Fig. 6.1 Off-tracking calculation process from visual odometry data

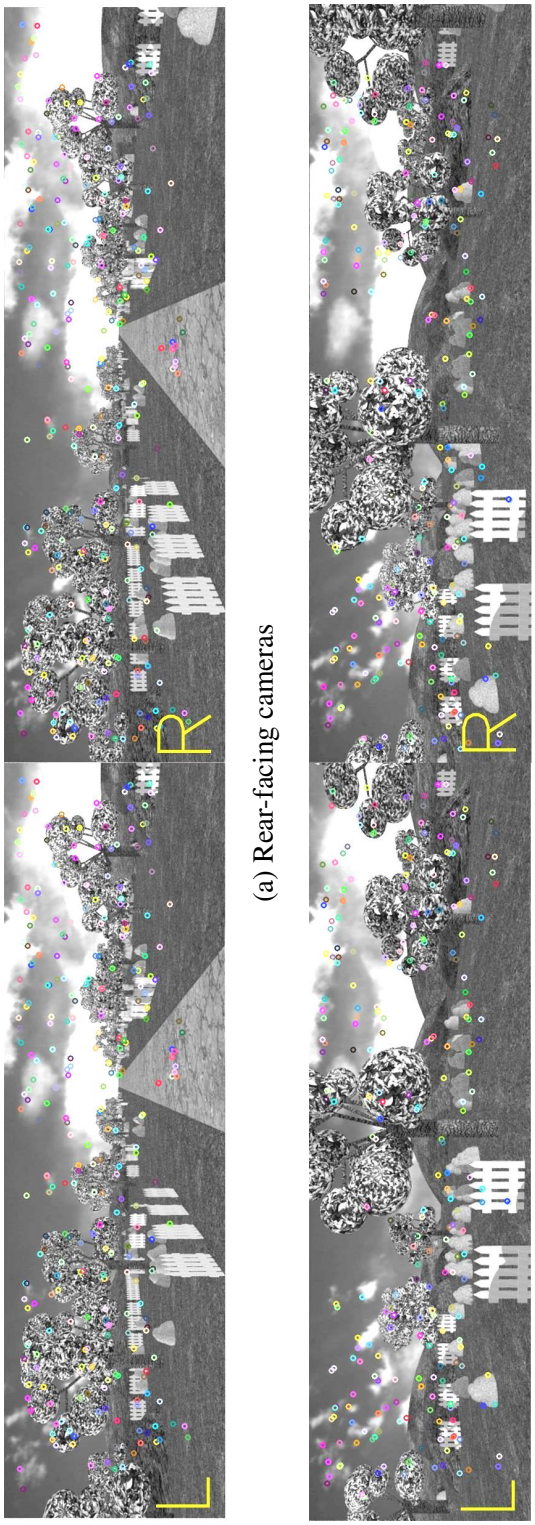


(a)



(b)

Fig. 6.2 CAD simulation environment: (a) Perspective view, (b) plan view of constant sideslip manoeuvre



(a) Rear-facing cameras

(b) Side-facing cameras

Fig. 6.3 Stereo images at a slip angle of 10° and a baseline of 2.5 m

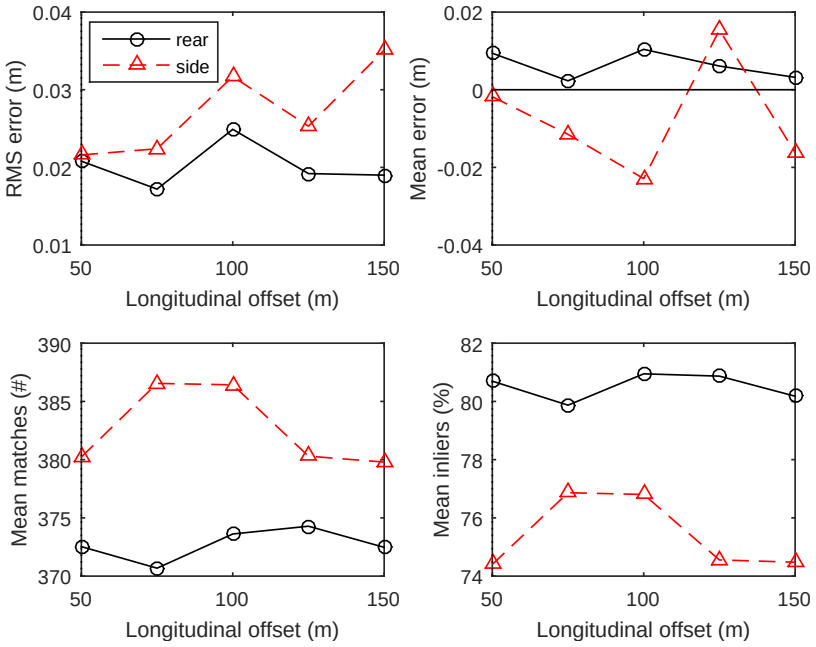


Fig. 6.4 Sensitivity to the *longitudinal offset*, measured from the camera starting point to the edge of the CAD environment (reference = 50 m)

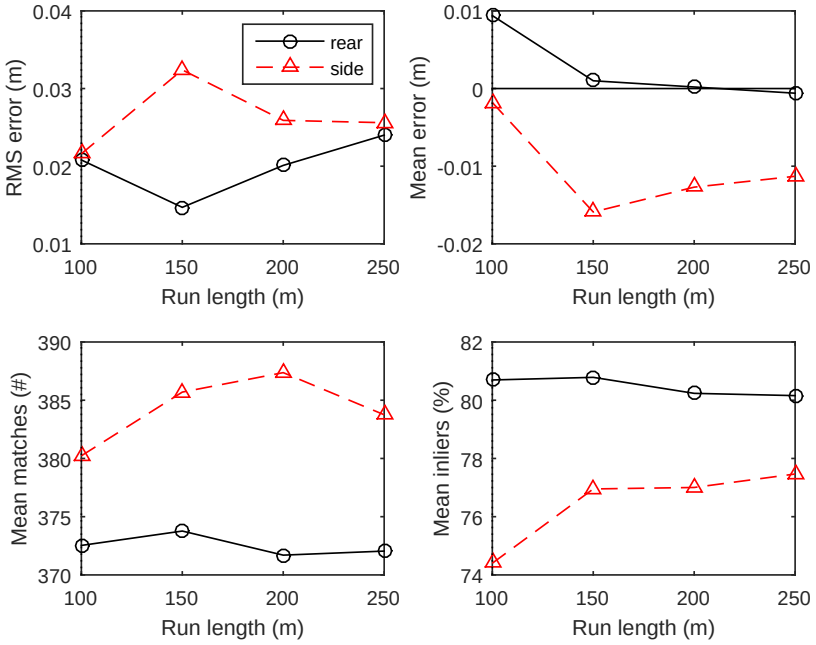


Fig. 6.5 Sensitivity to the *length of the simulation run* (reference = 100 m)

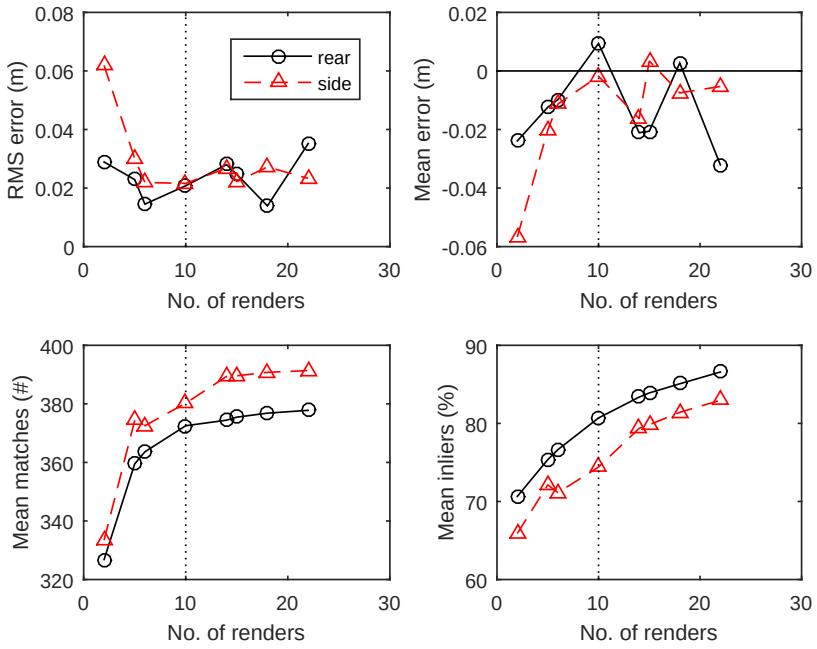


Fig. 6.6 Sensitivity to the *number of image renders* before processing (reference = 10)

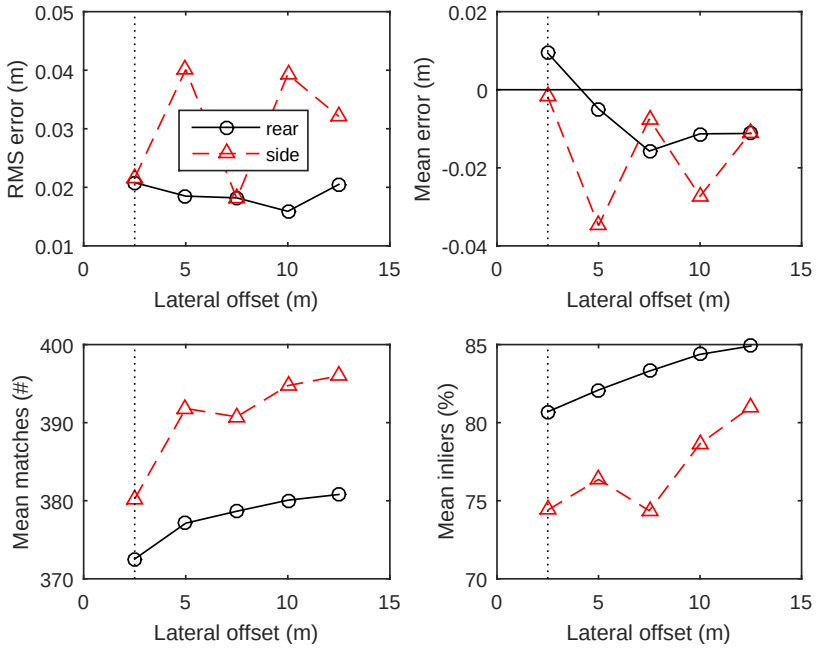


Fig. 6.7 Sensitivity to the *roadside lateral offset* (scenery distance from cameras) (reference = 2.5 m)

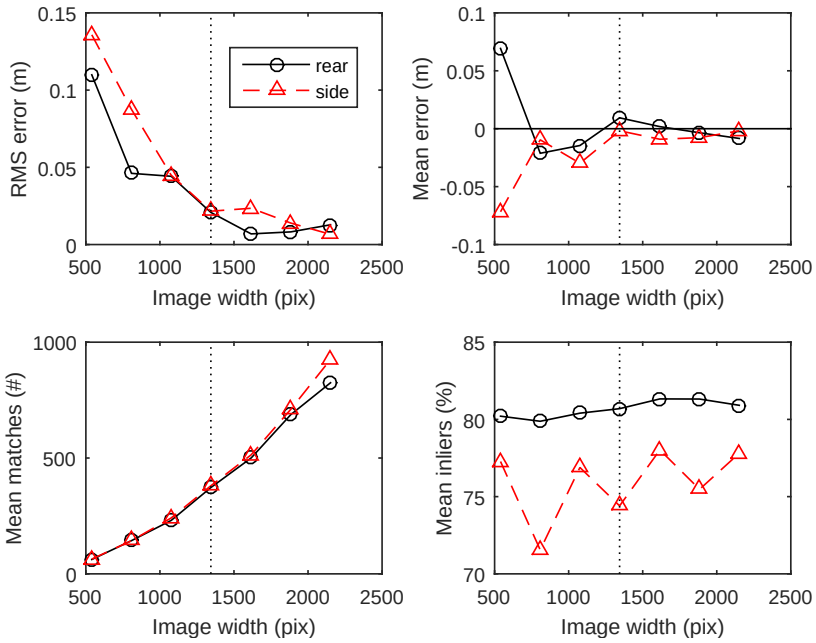


Fig. 6.8 Sensitivity to *image resolution* (as indicated by image width) (reference = 1344×391, aspect ratio fixed at 1:3.44)

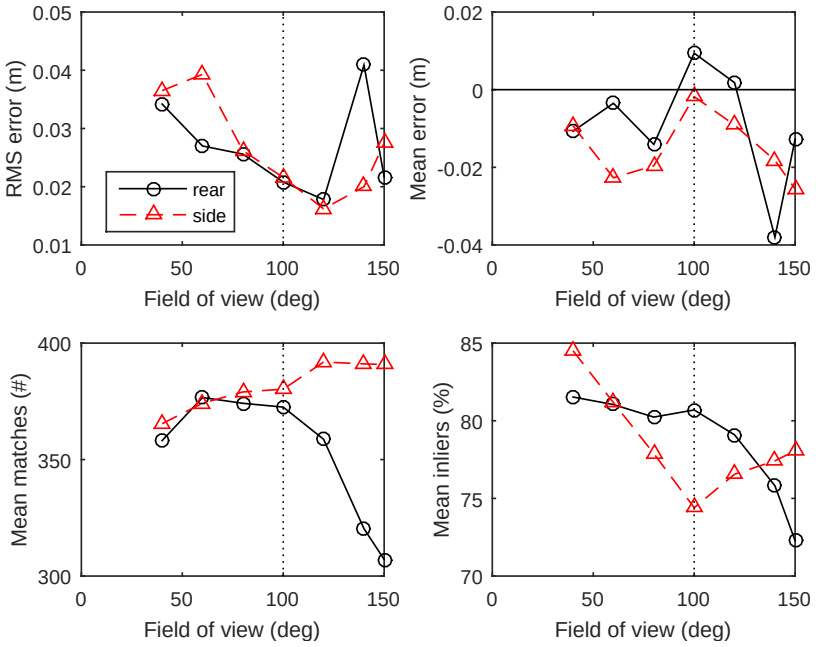


Fig. 6.9 Sensitivity to the camera *field of view* (reference = 100°)

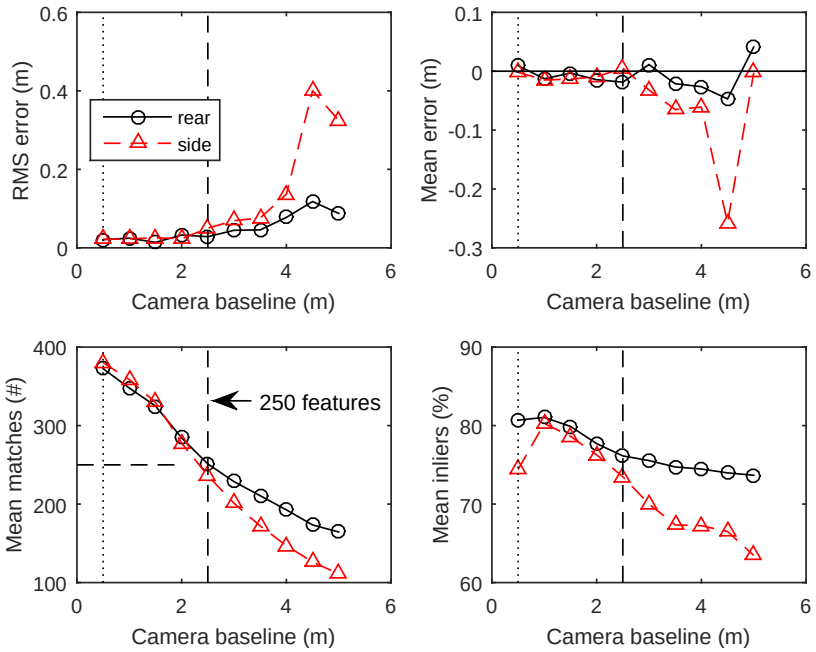


Fig. 6.10 Sensitivity to the *stereo baseline* for 14° slip (reference = 0.5 m)

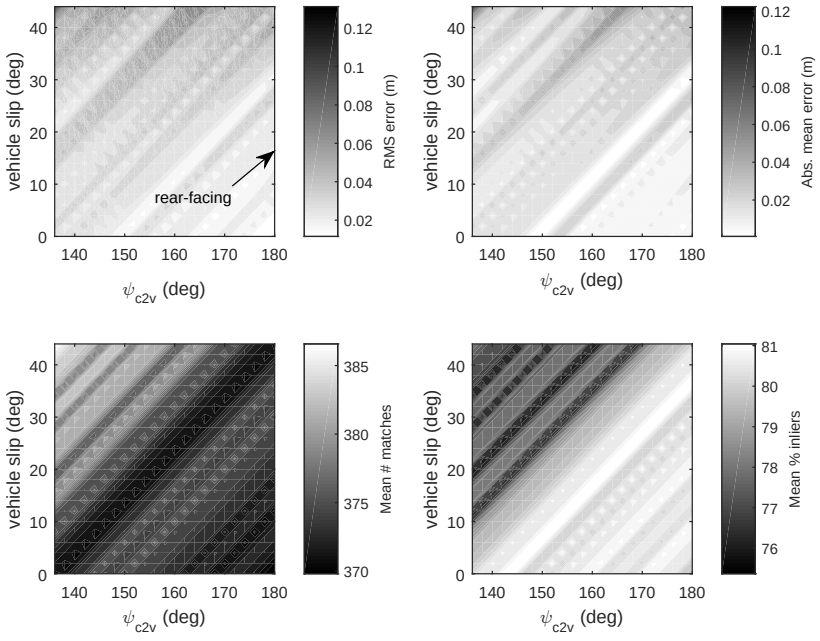


Fig. 6.11 Sensitivity to *camera rotation* relative to the trailer axis, as a function of trailer slip angle (180° = rear cameras, 90° = side cameras). Stereo baseline = 500 mm.

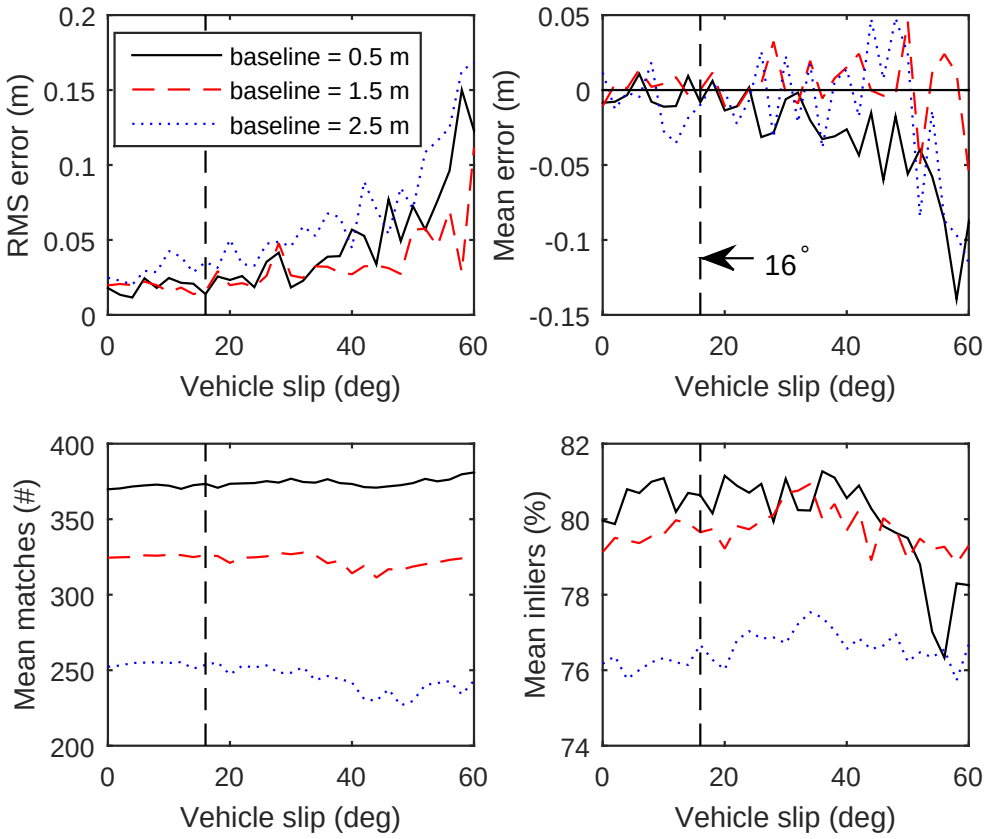


Fig. 6.12 Variation in performance for rear-facing cameras as a function of vehicle *slip angle and stereo baseline*. 16° trailer slip (4 m off-tracking) shown as dashed line.

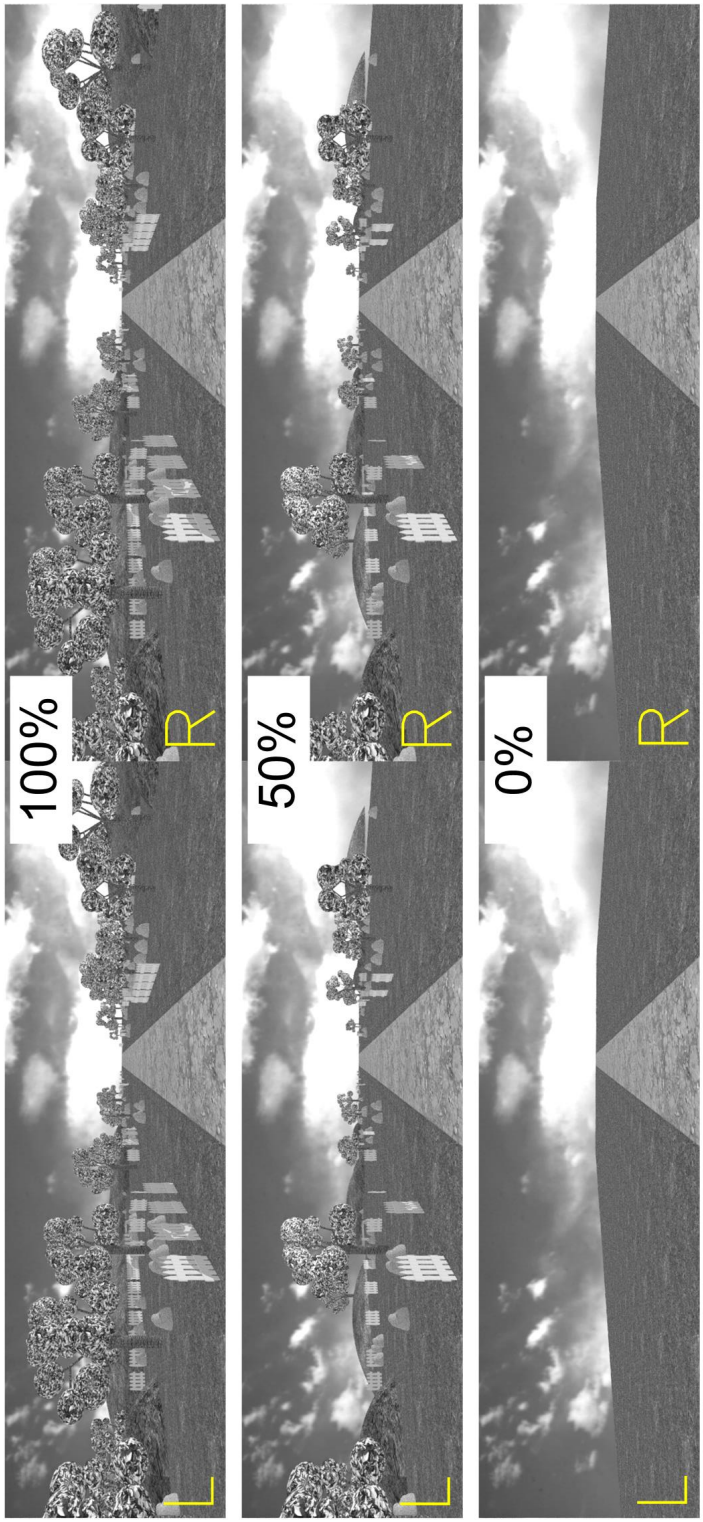


Fig. 6.13 Scenery variations of 100%, 50% and 0%, rear cameras

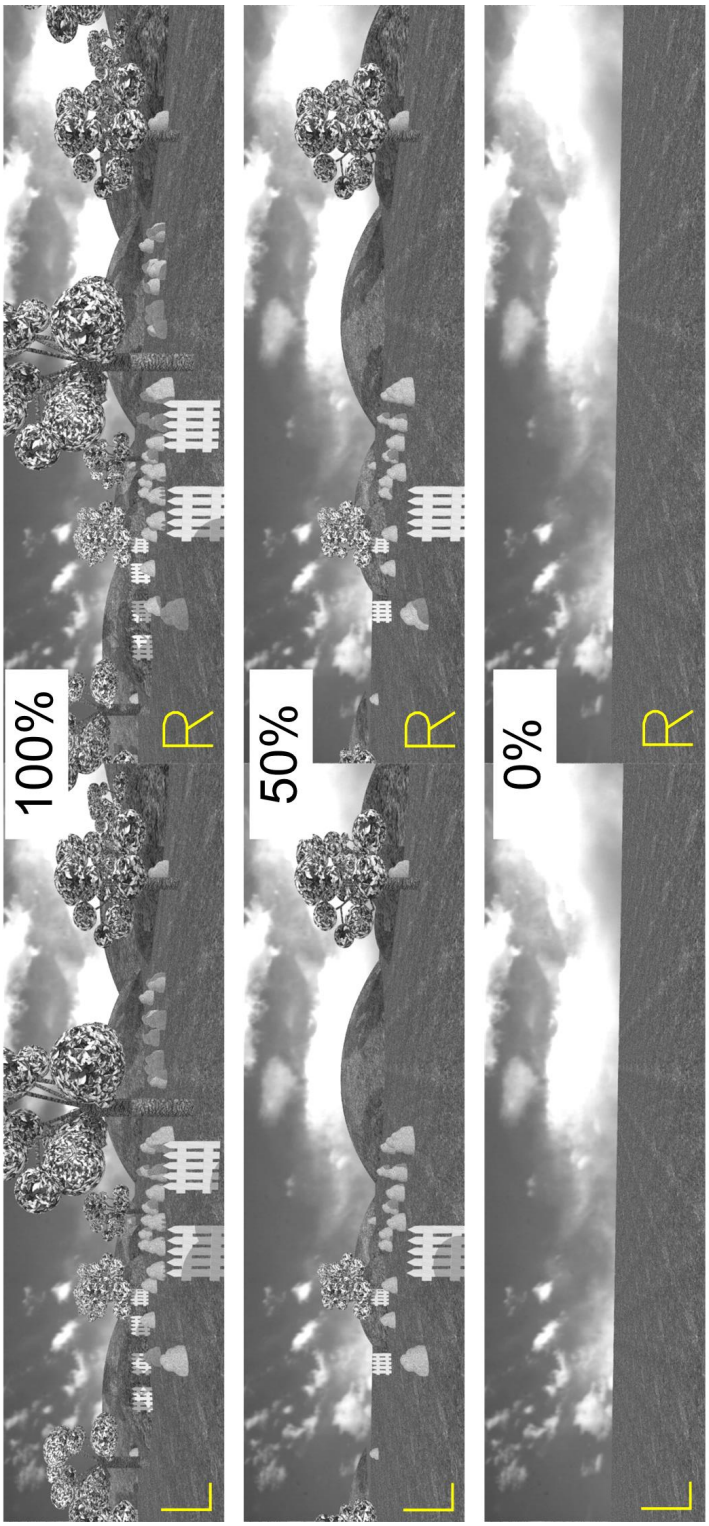


Fig. 6.14 Scenery variations of 100%, 50% and 0%, side cameras

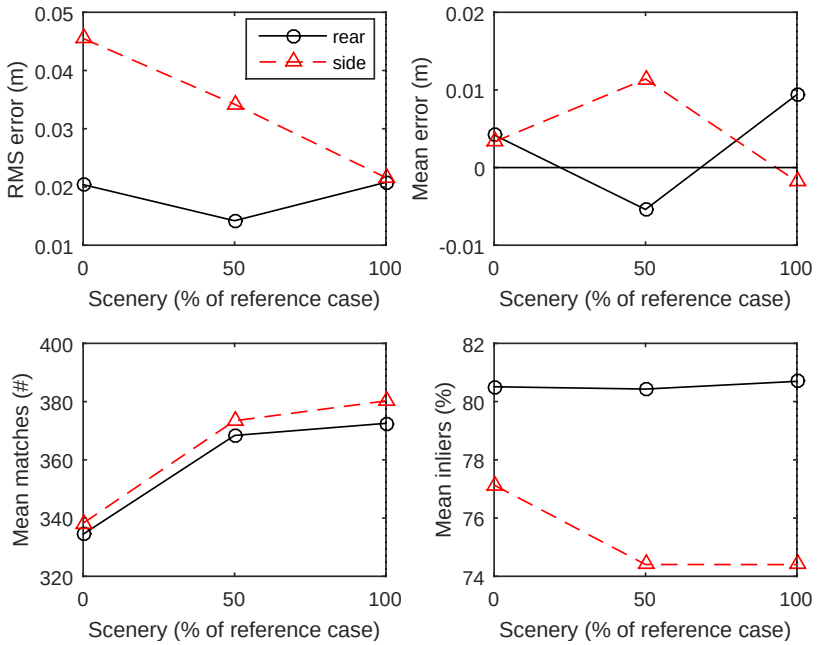


Fig. 6.15 Sensitivity to the density of the *roadside scenery* (reference = 100%)

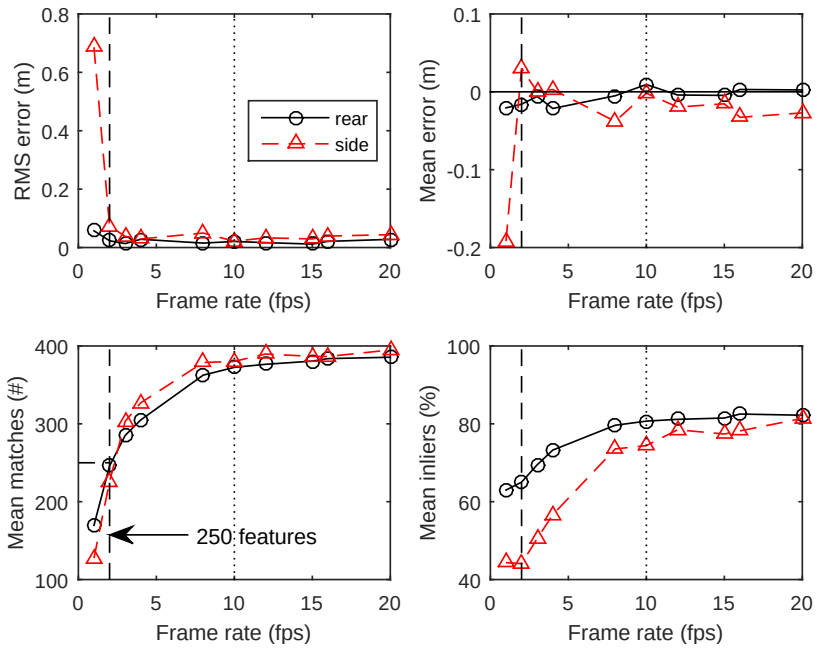
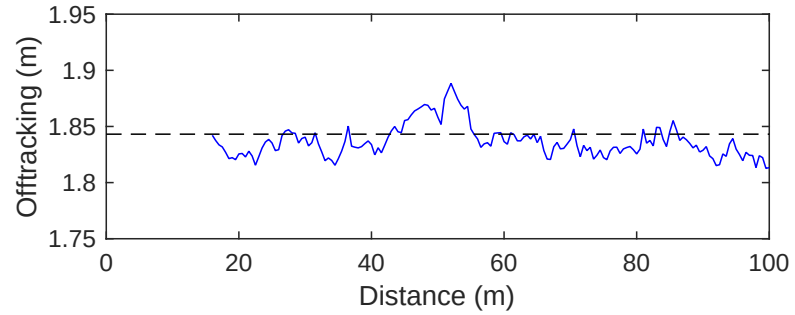
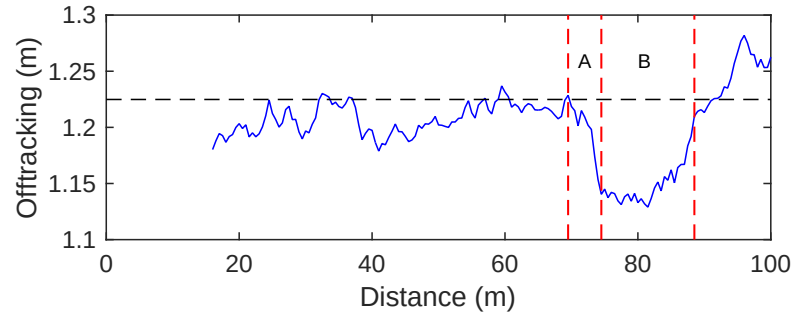


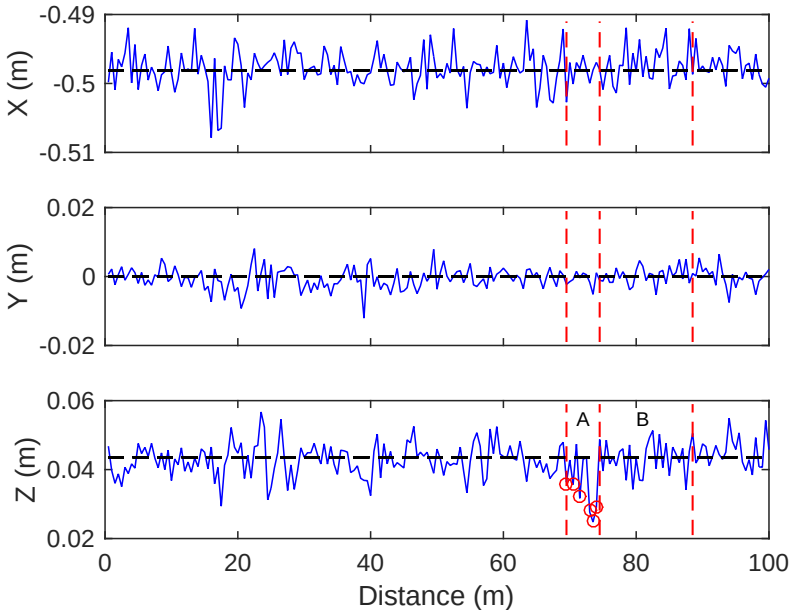
Fig. 6.16 Sensitivity to the *frame rate* (or inversely, vehicle speed) (reference = 10 fps)



(a)



(b)



(c)

Fig. 6.17 (a) Rear camera off-tracking results, 7.5° slip, 2.5 m baseline; (b) side camera off-tracking results, 5° slip, 1.5 m baseline; (c) visual odometry data for (b). (The camera optical axis is along the Z_c-axis, which is in the direction of off-tracking measurement for side-facing cameras.)

Chapter 7

Field tests of the trailer off-tracking sensor on a tractor semi-trailer

The previous chapter demonstrated a proof-of-concept for trailer off-tracking measurement using stereo cameras, and determined a limit on the expected measurement accuracy through simulation. The simulations were idealised in a number of ways, including: perfectly calibrated cameras with no distortion, consistent lighting conditions, and smooth travel of the cameras/trailer. In this chapter, the VISO-OT concept is implemented on a full-scale tractor semi-trailer combination and its real-world performance assessed through a series of manoeuvres in a visually-representative driving environment. Performance is compared to measurements obtained from a high-precision GPS navigation system.

7.1 Experimental setup

7.1.1 Vehicle and instrumentation

A full-scale tractor semi-trailer combination was used for testing (Figure 7.1), consisting of a 3-axle semi-trailer[1] (the same as in Chapter 4) and a Volvo FH12 6×2 tractor (different to Chapter 4). The tag axle of the tractor was permanently lifted during testing, as well as the front axle of the semi-trailer, making the combination effectively a 4×2 tractor with a 2-axle semi-trailer.

Two Point Grey Flea3 USB 3.0 cameras with wide-angle lenses were used for the stereo visual odometry, mounted to a custom machined aluminium mounting frame. The mount allowed for multiple camera baselines to be used, from 10 cm up to 100 cm in increments of

[1] Strictly speaking the trailer is a B-link trailer. However, for the purposes of this work the trailer is conceptually identical to a semi-trailer

10 cm, and was mounted to the top of the semi-trailer using strong magnets. The magnets allowed for easy removal of the mount to locate it at the rear or side of the trailer, or to adjust the baseline of the cameras. When mounted to the top of the semi-trailer, the cameras were approximately 4.2 m above ground. The mounting locations of the cameras are shown in Figure 7.1, and the stereo camera mount is shown in Figure 7.2.

The Point Grey cameras are the same as in Chapter 4 and Chapter 5, however new 2.9 mm fixed focal length lenses were acquired for these tests[2]. The field of view and focal range of these lenses were similar to those used previously. The cameras were mounted as as near to horizontal as possible.

The instrumentation and communication systems are shown in Figure 7.3. The same Linux computer from previous tests was mounted inside the trailer for image capture and processing, and connected to the cameras via USB 3.0. A trailer-mounted RT3022 inertial and GPS navigation unit was used to obtain measurements of global trailer position and heading angle, from which a 'ground truth' measurement of off-tracking was calculated[3]. The RT3022 contains precision accelerometers, gyroscopes, and GPS receivers, with dual GPS antennae to improve heading measurement accuracy. A GPS base station (the 'RT GPS-Base-2') was used to provide differential corrections, improving position measurements.

The RT3022 unit was rigidly mounted to the floor inside the trailer, aligned with the longitudinal axis of the trailer but offset laterally from the centre line due to practical constraints. The primary and secondary antennae were mounted on the roof of the trailer with a longitudinal separation of 2.342 m, also limited by practical constraints. The base station was located on an area of road adjacent to the test area (within 1 km of the test vehicle at all times), and placed with a good view of the open sky.

The RT3022 was configured to output selected measurements via CANbus, including heading and position, estimated heading and position measurement accuracies, and a number of status messages to monitor the operating modes and base station communication. These messages were transmitted through a PCAN-USB adapter (from PEAK-System Technik GmbH) to the Linux computer.

A laptop computer in the cabin of the tractor was connected to the image processing unit via ethernet. This was used to remotely access the Linux computer, to send commands to start and stop test runs, and to change camera and run parameters as needed.

[2]One of the original lenses had developed mechanical play, causing it to lose its focus setting during driving. Identical lenses could not be sourced at the time, and so comparable generic lenses were sourced for these tests.

[3]Off-tracking could also be calculated using a tractor-mounted RT3022 and articulation angle measurements from the VSE sensor. However, measurements of trailer position and heading directly from the trailer-mounted RT3022 yielded sufficient information for accurate off-tracking measurement using less instrumentation.

7.1.2 Test scenarios

Tests were conducted at Bourn Airfield near Cambridge. Visually, the test site consisted of an area of tarmac 40 m wide with various imperfections and potholes, surrounded by large fields of grain. An aerial view of the site is shown in Figure 7.4, including sample images of the surrounding scenery from the trailer-mounted cameras.

With the exception of the road surface which presented some helpful visual discontinuities, the grain fields surrounding the airfield provided a challenging scenario for the visual odometry algorithm. This represents the type of scenery to expect in a challenging off-highway application, and so provides a good assessment of the system's robustness.

Three types of manoeuvre were conducted, namely:

1. 'figures-of-eight',
2. left roundabout turns (three full turns per manoeuvre), and
3. right roundabout turns (three full turns per manoeuvre).

Rear camera ($\psi_{c2v} = 180^\circ$) and side camera (left-facing, $\psi_{c2v} = 270^\circ$) configurations were assessed at baselines of 500, 700 and 900 mm. Manoeuvres were conducted at relatively low speeds, in the range of 15 km/h (during cornering) to 40 km/h (during some straight sections). Potholes and other disturbances were not avoided by the driver, so as to include these disturbances in the assessment of the system's performance. Two of each manoeuvre were carried out for each configuration.

Before testing, the RT3022 was initialised according to the procedure prescribed by the manufacturer. The RT3022's reported measurement accuracies were monitored until they had reached acceptable stable values. This took approximately 15–30 minutes of driving above 5 m/s, including both straight and turning manoeuvres.

Before each test, the camera shutter speed and sensitivity was adjusted to suit the prevailing lighting conditions. A maximum shutter time of 2 ms was used to avoid motion blur, though often much lower than this in bright lighting conditions. When the image was too dark, additional image sensitivity gain was used to produce acceptable images while maintaining the maximum 2 ms shutter speed.

7.2 VISO-OT implementation

The VISO-OT algorithm was compiled from C++ on the Linux computer for real-time image processing. The computer uses a 3.4 GHz Intel Core i7 processor (full specifications can be found in Appendix A). The underlying VISO2-S and off-tracking algorithms were as

described in the previous chapter, but some additional functionality was added to the software system for the vehicle tests:

1. The off-tracking calculations were updated to account for laterally-offset cameras.
2. Real-time distortion correction and stereo rectification (Section 7.3.1) were added.
3. Off-tracking calculations and data logging for the RT3022 measurements were added.
4. Adjustable camera variables were introduced for shutter speed, gain, resolution, and region of interest. (These could be updated before each test.)
5. CANbus communication and data-logging functionality were added for RT3022 measurements (utilising the PCAN-Basic library for C++).
6. Additional user interface and data visualisation functionality was added for testing and troubleshooting (utilising the OpenCV C++ library).

The underlying VISO2-S algorithm was used in much the same configuration as was used in [133]. This includes two-scale feature matching, 50×50 bucketing windows and 50 RANSAC iterations. Ten features per bucket were adopted here (instead of two as before), as this improved accuracy and was able to maintain 10 fps. An image resolution of 1328×388 was used, giving a wide aspect ratio which helped provide high resolution for motion in the yaw plane while reducing processing demands by not using the entire field of view.

As the native camera resolution was 1328×1048, the narrower resolution was obtained by defining a region of interest within the image. The region of interest was chosen to be vertically offset from the centre of the native image field of view. This offset is denoted v-shift, and is shown in Figure 7.5. An offset of 120 pixels downwards was used for the rear cameras, ensuring a visible horizon, while minimising extraneous sky detail in the image to maximise the number of useful features. A vertical shift of 70 pixels was used for the side cameras to compensate for a small change in the pitch of the cameras. This was a result of imperfections in the trailer structure where the stereo cameras mount was attached.

The updated VISO-OT algorithm was able to run at approximately 10 fps. This included image capture, distortion correction, stereo rectification and processing. Faster processing of up to 20 fps was possible by reducing the number of features per bucketing window by half, resulting in a small loss of accuracy. For these tests the full number of features was adopted, and hence processing was limited to 10 fps.

Equation 6.2 described how displacements of the fifth wheel per image frame ($\Delta x_{F,i}$, $\Delta y_{F,i}$) could be obtained from visual odometry data measured at the location of the cameras (Δx_i, Δy_i). This was obtained using the change in yaw angle, $\Delta \psi_i$, and knowledge of the camera location in the form of parameter a.

Equation 6.2 is true if there is no lateral offset between the cameras and the longitudinal axis of the trailer. For the vehicle tests there was a lateral offset which had to be accounted for. Equation 6.2 can be updated for the current application as follows:

$$\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}_{F,i} = \begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix}_{i} + \begin{bmatrix}\cos(\Delta\psi_i)-1 & \sin(\Delta\psi_i)\\ -\sin(\Delta\psi_i) & \cos(\Delta\psi_i)-1\end{bmatrix}\begin{bmatrix}c\\ a\end{bmatrix} \tag{7.1}$$

where parameters a, b and c are defined in Figure 7.6, and measurements of a, b and c used for testing are shown in Figure 7.7.

Referring back to Figures 7.4 and 7.5, consider briefly the effect which moving objects might have on the visual odometry measurements, given that these would contribute biasing feature points in the temporal feature-matching step. A scale is shown in Figure 7.4 for reference. Individual passenger vehicles (approximately 2 m wide, 1.5 m tall, 5 m long) would occupy up to possibly 10% of the rear images and 20% of the side. Through the bucketing and RANSAC processes, odometry measurements should be adequately isolated from these biasing feature points; multiple passenger vehicles may have more of an impact. A tractor semi-trailer (approximately 2.5 m wide, 4 m tall, 16.5 m long) would occupy up to possibly 25% of the rear view which should not be problematic, but up to possibly 50% of the side which could have a significant biasing effect. However, instances of a passing tractor semi-trailer, or where multiple passenger vehicles are in view at once, are likely to only occur on primary roads and motorways, and often only during straight road sections. In these conditions the trailer steering system is unlikely to be in operation, and so the VISO-OT concept is expected to be robust to moving vehicles in most operating conditions.

7.3 Stereo camera setup

7.3.1 Stereo calibration

In previous chapters, a camera calibration process was required to determine the ‘intrinsic’ camera parameters—the camera matrix and distortion coefficients—in order to ‘undistort’ images (Section 2.3). For the stereo camera rig, an additional ‘stereo calibration’ was required to determine the ‘extrinsic’ camera parameters—the relative orientation between left and right cameras—in order to ‘rectify’ the stereo images (Section 2.4).

The VISO2-S algorithm requires knowledge of the focal length f and optical centre (u_0,u_0) for each camera, as well as the stereo baseline. Additionally, the algorithm assumes that input images are undistorted and rectified.

A C++ script was developed for intrinsic and extrinsic camera calibration, making use of the FlyCapture and OpenCV C++ libraries for camera interfacing and calibration functions respectively. Left and right intrinsic camera calibrations were performed at the full resolution of 1328×1048, using the same checkerboard pattern as before (Chapters 4 and 5), with a 7×10 grid and a square size of 25.5 mm. Approximately 30 images were used per calibration, and reprojection errors of less than 0.5 pixels were achieved for both cameras.

After this, stereo calibration was performed at full 1328×1048 resolution, using a larger 7×10 checkerboard with a square size of 93.33 mm. The larger checkerboard was necessary given the relatively large stereo baselines used, so that the full checkerboard was visible in both left and right images for a suitable range of viewpoints, while also ensuring that the board occupied a significant portion of the pixel space in each image. An example stereo calibration image pair of the large checkerboard pattern is shown in Figure 7.8.

Approximately 30 images were used per stereo calibration (for baselines of 500, 700 and 900 mm), yielding reprojection errors close to 0.5 pixels in all cases. Due to practical limitations, calibration was carried out before the stereo camera mount was attached to the trailer. Intrinsic and extrinsic calibration results are given in Appendix C.

Distortion correction and stereo rectification were performed in real time, incorporated into the processing loop of the visual odometry algorithm in C++ using OpenCV's 'undistort()' and 'stereoRectify()' functions.

7.3.2 Stereo camera triggering

Ideally, stereo cameras should be triggered in perfect synchronisation. This can be achieved using an external trigger, or may be approximated by triggering the cameras via software, provided the delay between the two triggers is small relative to the baseline divided by the speed of travel.

The Flea3 cameras have an on-board buffer, which allows them to be triggered independently of retrieving images. This allows for triggers to be sent in quick succession to left and right cameras without waiting for images to be transferred, limited only by code execution time. The images may then be retrieved after both cameras have been triggered.

The cameras were triggered sequentially in the loop of the C++ program, and stereo trigger delays of between 0.1 and 0.3 ms were observed. At 10 fps, with a vehicle speed of 5 m/s, the maximum 0.3 ms stereo delay translates into a distance of 1.5 mm. This is either 0.3% difference in effective baseline for side-facing cameras (for a 500 mm baseline), or 0.3% difference in distance travelled per frame for rear cameras. This delay was deemed suitable relative to the best-case accuracy observed in the CAD tests, coupled with additional expected errors due to camera calibration and optical noise expected during field testing.

7.4 RT3022 'ground truth' measurements

In previous work [28, 32], ground truth off-tracking measurements were obtained using two downwards-facing cameras which tracked a painted line on the road along which the vehicle was driven. One camera was placed beneath the fifth wheel and the other below the rear follow point on the trailer. The relative position of the line in the images between front and rear cameras was used to calculate off-tracking.

Subsequently, Rimmer [47] used an RT3022 and articulation angle measurements to calculate the off-tracking of multiple trailers. This was implemented alongside the above camera-based approach, and found to provide comparable results without the requirement for a marked path. Articulation measurements were also required as multiple trailers were used.

In this work, RT3022 measurements alone were used to provide ground truth off-tracking measurements. This was achieved using trailer position and heading measurements from the RT3022. The method used to calculate off-tracking from heading and position measurements is now described.

The global position of the fifth wheel ($x_{i,F}$, $y_{i,F}$) and the follow point ($x_{i,R}$, $y_{i,R}$) can be calculated as follows:

$$\begin{bmatrix} x \\ y \end{bmatrix}_{F,i} = \begin{bmatrix} x \\ y \end{bmatrix}_{RT,i} + \begin{bmatrix} \cos(\Psi_i) & \sin(\Psi_i) \\ -\sin(\Psi_i) & \cos(\Psi_i) \end{bmatrix} \begin{bmatrix} a_{RT} \\ c_{RT} \end{bmatrix} \tag{7.2}$$

$$\begin{bmatrix} x \\ y \end{bmatrix}_{R,i} = \begin{bmatrix} x \\ y \end{bmatrix}_{RT,i} + \begin{bmatrix} \cos(\Psi_i) & \sin(\Psi_i) \\ -\sin(\Psi_i) & \cos(\Psi_i) \end{bmatrix} \begin{bmatrix} -b_{RT} \\ c_{RT} \end{bmatrix} \tag{7.3}$$

where Ψ is the RT3022 heading angle, relative to the North direction and positive clockwise, and a_{RT}, b_{RT}, and c_{RT} are the locations of the RT3022 relative to the trailer body (Figure 7.6). Figure 7.7 shows the measured values of a_{RT}, b_{RT}, and c_{RT} used in testing.

Using a data buffer in a similar manner to Section 6.1.2, the global (x, y) co-ordinates of the fifth wheel trajectory, relative to the i-th fifth wheel location, may be found as follows:

$$\begin{bmatrix} x \\ y \end{bmatrix}^{(i)}_{F,i-k} = \begin{bmatrix} \cos(\Psi_i - \frac{\pi}{2}) & \sin(\Psi_i - \frac{\pi}{2}) \\ -\sin(\Psi_i - \frac{\pi}{2}) & \cos(\Psi_i - \frac{\pi}{2}) \end{bmatrix} \begin{bmatrix} x_{F,i-k} - x_{F,i} \\ y_{F,i-k} - y_{F,i} \end{bmatrix} \tag{7.4}$$

where k is the buffer counter and the superscript (i) refers to the i-th reference frame as before. Off-tracking may then be calculated by finding the point in the buffer at which $y^{(i)}_{F,i-k} = -(a+b)$, and using Equation 6.7.

When used with only a single antenna, the RT3022 is able to provide heading measurements to an accuracy of 0.1° RMS and can result in significant heading drift when stationary. Heading precision and drift can be significantly improved by utilising the secondary antenna, mounted a fixed distance behind the primary antenna. When used in dual antenna mode, with a 4 m antenna separation, a quoted RT3022 heading precision of 0.05° RMS can be obtained.

The RT3022 yields GPS position measurements with a quoted accuracy of 1.5 m CEP (circular error probable) when no external corrections are used. Position accuracy can be improved by incorporating differential GPS corrections through the use of a GPS base station. The RT GPS-Base-2 was used which supplies L1/L2 corrections[4], reducing position errors to approximately 0.02 m RMS. The base station communicates with the vehicle-based RT3022 via radio.

These RT3022 off-tracking calculations incorporate summations and hence it is possible for errors to accumulate over the length of the trailer. However, this was deemed satisfactory given the high precision of the RT3022 measurements and assuming errors to be zero-mean Gaussian. Assuming the quoted 0.05° RMS heading error and 0.02 m RMS position error, and using the fifth-wheel-to-follow-point distance of 11.5 m (see Figure 7.1), the maximum expected RMS off-tracking error can be approximated as $0.02 + 11.5 \times \sin(0.05^\circ) = 0.03$ m, which is sufficiently small relative to the target accuracy (maximum and RMS errors of 0.15 m and 0.10 m respectively).

The quoted position and heading accuracies of the RT3022 are upper bounds, and the actual precision can be expected to vary depending on a number of factors. These factors include the accuracy with which the RT3022 and antenna locations were measured and input to the RT3022, the visibility of satellites by both vehicle-based antennae and the base station, line-of-site of the radio aerials, and whether or not the system was successfully initialised. The RT3022 provides internally calculated estimates of its real-time measurement accuracies using a Kalman Filter and user-provided uncertainties in the RT3022 and antennae mounting locations. These estimates were logged via CANbus during testing to give a better indication of ground truth measurement accuracies during testing.

7.5 Bias correction

Bias in off-tracking measurements from the stereo cameras or the RT3022 could arise from trailer crabbing, misalignment of the cameras/RT3022 relative to the trailer, or misalignment

[4]GPS signals are affected by the density and make-up of the atmosphere, and so accuracy can be affected by varying conditions. Corrections for this can be made by utilising two different carrier frequencies, L1 and L2. By comparing the signal delays of both, these errors may be accounted for, giving better precision GPS measurements.

of the trailer body relative to the trailer chassis. This was accounted for by performing straight line driving manoeuvres before each set of manoeuvres, and after any changes to the stereo camera baseline or mounting location.

Denoting the measured bias in stereo camera off-tracking measurements as $e_{tr,\,\text{bias}}$, bias correction was incorporated into the ψ_{c2v} parameter in Equation 6.1. The original ψ_{c2v} parameter was defined according to the chosen camera mounting orientation (rear or side), and is redefined here as $\psi_{c2v,\,\text{rear/side}}$. The bias-corrected value of ψ_{c2v} was then calculated as follows:

$$\psi_{c2v} = \psi_{c2v,\,\text{rear/side}} + \arctan\left(\frac{e_{tr,\,\text{bias}}}{a+b}\right) \tag{7.5}$$

Correcting RT3022 bias was performed in a similar manner, by correcting heading measurements Ψ_i in Equations 7.2, 7.3 and 7.4.

Two straight line driving manoeuvres were conducted per camera configuration (left/side, baseline changes), and the average bias used for correction.

7.6 Results

7.6.1 Bias correction

Biases in the RT3022 off-tracking measurements during the straight-line driving manoeuvres were measured to be in the region of 0.007–0.030 m. These values could have been a result of a small misalignment of the RT3022 mounting in the trailer, crabbing behaviour of the trailer due to axle misalignment, misalignment of the two GPS antennae, camber in the road surface, or some combination of the above. The magnitude of the biases were small enough to rule out any major crabbing or misalignment problems, and were accounted for in the processing of all results.

Biases in the stereo camera off-tracking measurements during the straight-line manoeuvres were in the region of 0.1–0.3 m for the rear camera tests and 0.48–0.96 m for the side cameras, indicating a likely misalignment in the camera mount relative to the direction of travel. For both side/rear mounting configurations and for all stereo baselines, the camera mount was aligned with the edge of the trailer as closely as possible. However, the side camera mounting was particularly challenging due to structural constraints.

It was estimated that the accuracy of this alignment with the trailer edge was approximately 1°. However, over 11.5 m from fifth wheel to follow-point, an observed 0.3 m bias would suggest a total of 1.5° misalignment, and 0.96 m suggests 4.8° of misalignment.

Therefore another component of misalignment must exist, and it was concluded that this was probably a result of structural misalignment of the top of the trailer relative to the chassis.

The lack of a large bias in the chassis-mounted RT3022 suggested that trailer crabbing (due to wheel alignment errors) was not an underlying cause of biases. However, the described method for accounting for the biases is independent of the source of bias, and would still provide a like-for-like comparison between the camera and RT3022 measurements.

7.6.2 Ground truth errors

RMS errors reported from the RT3022 during testing averaged in the region of 0.03° heading and 0.04 m position. Using the simple relationship given before, this gives approximately 0.05 m RMS off-tracking error. This is slightly above the original estimate of 0.03 m (Section 7.4), but is still acceptable relative to the target precision of the visual odometry measurement.

7.6.3 Camera mounting and initial feature matching quality

Example stereo image pairs overlaid with successful feature matches are shown in Figure 7.9. The 900 mm baseline cases are shown at the beginning of the figure-of-eight manoeuvre, selected at approximately the same location. The distribution of features evident in these examples is generally representative of all tests.

In some cases, a notable change in the distribution of features was evident before and after mounting the cameras to the trailer. In some cases, feature matching would fail entirely after the camera mount was attached to the trailer. It was supposed that this was due to local deformations in the trailer roof. Because of the strength of the magnets at the centre and left and right ends of the camera mount, mounting the frame caused it to contort to the profile of the trailer roof, including where these deformations were present.

For practical reasons the cameras were calibrated at ground level, with the mount unattached. Any bending or twisting of the camera mount described above would result in a change to the relative orientation of the left and right cameras, hence affecting the validity of the stereo calibration parameters. The VISO2-S algorithm assumes the stereo images to be fully rectified in order to reduce the stereo matching process to a simple horizontal scanning task (see Section 2.4), and if the cameras are twisted sufficiently out of calibration, this matching is expected to fail.

Where the effect of this was pronounced, it was remedied using small spacers between the camera mount and the trailer roof, until visually acceptable feature matching quality was achieved (Figure 7.9 shows an example of ‘acceptable’ feature matching quality).

7.6.4 Off-tracking measurements

Off-tracking measurements as a function of time are given in Figures 7.10, 7.11, 7.12, for the 500, 700 and 900 mm stereo baselines respectively. In each figure, results are shown for all three manoeuvres and for both rear and side camera configurations. Results for one of the two tests per manoeuvre are shown. Included beneath each plot is the time history of the number of successful feature matches per image pair (after outlier rejection), as well as the calculated error signal between the VISO-OT and ground truth measurements.

At the beginning of each plot there is a period of between 50 and 100 frames without data (≈5–10 s). This is the period during which one trailer length of data is being accumulated in the buffer (for both VISO-OT and RT3022 results), and so off-tracking is not yet defined. These data were manually removed so as to not affect overall error metrics for each test run.

Overall, off-tracking measurement accuracy is reasonable, with errors generally less than 0.5 m. However, errors of up to approximately 1 m are evident in certain cases, including Figure 7.11d and Figure 7.12e. Large oscillations in off-tracking measurement are clear in some cases (Figure 7.11d), and there is evidence of significant bias in others (all right roundabout manoeuvres, Figures 7.10e, 7.10f, 7.11e, 7.11f, 7.12e, 7.12f). The number of features matched per image pair is in the region of 400–800 features in most manoeuvres, but as low as 50 features in one instance (Figure 7.10c). These observations will now be discussed in more detail.

The number of feature matches is well above the 250 threshold identified in the previous chapter, below which errors were expected to rise steadily. The number of matches drops to about 200 features in some cases, without an observed effect on errors. In Figure 7.10c however, the number of feature matches drops to well below this to approximately 50 features, although there is no evidence of this having affected the measurement errors.

There is a clear periodic element to the number of feature matches in the left and right roundabout results. There appear to be either three or six cycles per manoeuvre, which is consistent with the number of turns driven per manoeuvre. At the time of testing the sun was setting in the west, and the variations in feature matching can probably be attributed to variations in lighting conditions as the cameras faced towards and away from direct sunlight. These oscillations in feature matches do not seem to have affected any measurement errors directly, as the total number of features matches has remained acceptable. These effects could be reduced in future work through the use of more advanced automatic shutter speed and exposure adjustment, compared to the manual adjustment adopted for these tests.

Large oscillations in off-tracking measurement and hence error during the left roundabout manoeuvre are evident in Figure 7.11d, which appear to be correlated with variations in feature matches. However, maximum errors do not correlate with minimum features as

might be expected; in fact the opposite is observed. The oscillations exhibit approximately six cycles, which is consistent with the three 360° turns performed during the roundabout manoeuvre. However, it can be calculated that approximately 2.2 trailer lengths were travelled by the rear of the trailer during each turn, and so these variations could also be attributed to the filling and emptying of the data buffer (see Section 6.3.5 and Figure 6.17b). It is likely to be a result of a combination of these effects.

Figure 7.10d exhibits similar oscillations, but smaller in magnitude and with approximately twelve cycles. Although this cannot be attributable to sunlight variations alone, it is also consistent with the number of turns, with four variations per turn. The observed scenery during the manoeuvre can be categorised into four distinct parts per turn, and for this manoeuvre transitioned from grain fields/road in the north, to the westwards length of the runway, to the grain fields in the south, and finally to the eastwards length of the runway. An appropriate automatic method for adjusting camera parameters should help to reduce sensitivity to lighting and scene variations.

There is a recurring error trend in the rear-facing camera results, in that positive off-tracking ($+e_{tr}$) generally yields small errors, while off-tracking in the negative direction ($-e_{tr}$) yields a consistent bias. This is clear when comparing results for left and right roundabout turns: all right roundabout turn results exhibit a clear measurement bias, underestimating off-tracking by 0.5–1 m, whereas left roundabout turn results show no evidence of this. In results for the figure-of-eight manoeuvres, the largest errors are consistently in the region of negative off-tracking.

A seemingly similar trend is also evident in the side camera results, but is less consistent, and trends are more difficult to observe as the side-camera results are generally more erratic. Biases in the right roundabout manoeuvres are in the negative off-tracking direction for the 500 mm and 700 mm stereo baselines (Figures 7.10f and 7.11f), but in the positive off-tracking direction for the 900 mm baseline case (Figures 7.12f). These biases do not originate from zeroing bias, as it only appears at non-zero off-tracking magnitudes. Corrections from the straight-line driving manoeuvres have already accounted for zeroing bias.

For each test, RMS and maximum errors over the course of the manoeuvre were calculated. Feature matching statistics including the mean number of features matched, and the number of these which survived the outlier rejection process, were also recorded. These results are shown in Figure 7.13 for all three manoeuvres (averaged for the two test runs), for both side and rear camera configurations.

Although the results are erratic as a result of the afore-mentioned oscillations and bias, in general the rear cameras appear to provide better accuracy than the side cameras. The left roundabout tests (Figure 7.13b) were the least affected by bias (the cause of this is discussed

in the next section), and so provides a reasonable picture of the achieved accuracies. In these tests RMS and maximum errors of 0.11–0.12 m and 0.31–0.44 m were obtained for the rear cameras, and 0.30-0.59 m (RMS) and 0.36–1.16 m for the side cameras.

There seems to be an overall increase in errors with increasing baseline for all tests. However, there appears to be a small but convincing increase in feature matches with increasing baseline for both rear and side cameras. The scenery in these tests was closest to the '0%' scenery case studied in Chapter 6, in which average feature depth is at a maximum. In this case this should favour larger baselines as is observed. However, the effect is small, and the number of features is already well above the required minimum (≈ 250), and so there does not seem to be a benefit to baselines beyond 0.5 m.

The target RMS and maximum errors of 0.10 m and 0.15 m were not met, but the rear camera configuration came close. An improved stereo camera mount, automatic camera adjustment to lighting conditions are expected to reduce these errors to acceptable levels.

The consistency of the errors within each camera configuration for all three manoeuvres suggests an underlying systematic effect. It was hypothesised that these errors originated from imperfections in the trailer structure, causing small pitch and/or roll deviations between the left and right cameras. Further, small yaw angle deviations could have resulted where the cameras were attached to the mount during baseline changes. These would have led to small yet sufficient deviations from the stereo calibration (particularly R_{rl}, Section 2.4), so as to bias the visual odometry measurements in one direction or another, but not so large as to prevent feature matching entirely (observed during initial camera mounting).

The magnitude and direction of pitch, roll and yaw variations would be expected to differ between stereo baselines (as cameras are moved to different positions on the mount) and side/rear camera configurations (as the mount is moved between rear and side mounting locations). However, these are expected to be consistent for all three manoeuvres for a given camera configuration if this hypothesis is correct. Additionally, the trailer structure itself could deform a result of moments in the chassis generated during turning. This may have acted to correct for these camera rotations in left turns, but amplify them in right turns.

7.6.5 Camera rotation correction

To validate the camera rotation hypothesis, a small investigation was carried out in which artificial pitch, roll and yaw angle transformations were applied to right camera images in post-processing. This would 'correct' for the theorised rotation incurred by twisting or bending of the camera mount. If the rotation angles are chosen correctly, the systematic errors in observed off-tracking measurements should diminish.

If the right camera is assumed to have rotated about its optical axis, and if roll, yaw and pitch rotation angles are denoted ϕ, ψ and θ as before, a 'planar projective transformation' can be applied to the right camera images to simulate 3-D camera rotation as follows [157]:

$$\tilde{\mathbf{w}} = \mathbf{K}\mathrm{R}\mathbf{K}^{-1}\tilde{\mathbf{w}}_0 \tag{7.6}$$

where

$$\mathrm{R} = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{7.7}$$

and where $\mathbf{K}$ is the camera matrix, $\tilde{\mathbf{w}}_0$ is the image before camera rotation, and $\tilde{\mathbf{w}}$ is the image after camera rotation. This transformation does not assume a planar scene.

The three manoeuvres of the rear-facing 500 mm baseline case were used for a small parametric study. Transformations with different combinations of ϕ, ψ and θ were applied to the right camera images in each manoeuvre, image sequences were reprocessed, and the effects of the transformation on off-tracking results were observed.

A combination of $\phi = 0.15°$, $\psi = -0.10°$ and $\theta = 0°$ yielded a plausible and consistent set results for all three manoeuvres (suggesting that the right camera was rotated by $\phi = -0.15°$ and $\psi = 0.10°$ during testing). Results are shown in Figure 7.14 where the new results (right column) are shown alongside the original results (left column) for each manoeuvre. The angle corrections have corrected the bias for $-e_{tr}$ values, while not significantly altering errors for $+e_{tr}$.

These results are not without fault, such as the increased error at approximately 87 s in Figure 7.14f which seems to be the result of an induced lag effect. This could be the result of additional effects which have not been modelled such as translation offsets. However, these results, together with the observed visual odometry sensitivity during camera mounting, confirm that very small ($< 0.2°$) changes in the camera orientation can significantly affect off-tracking measurements, and this is likely to have been the case during these experiments.

Figure 7.15 shows a summary of the effects of these corrections on RMS and maximum errors. In general the effect of the correction is positive (with the exception of the maximum error in the right roundabout manoeuvre), and RMS errors are more repeatable between manoeuvres. Corrected RMS errors are in the range 0.11–0.13 m, close to the target of 0.1 m. This suggests that the target accuracy is likely to be attainable, should these experiments be repeated with an updated camera mounting solution.

7.7 Conclusions

1. The visual odometry-based off-tracking estimation concept was implemented on a tractor semi-trailer combination for field testing. Low-speed manoeuvres were performed at Bourn airfield on dry tarmac.

2. Side and rear facing camera configurations were investigated at stereo baselines of 500, 700 and 900 mm. The surrounding roadside scenery consisted predominantly of grain fields with few obvious visual features such as trees or buildings, providing a challenging visual scenario for assessment.

3. An RT3022 inertial and GPS navigation system was used to obtain ground truth measurements. Dual GPS antennae were used to improve heading accuracy, and a GPS base station was used to improve location accuracy.

4. In left roundabout manoeuvres, errors of 0.11–0.12 m (RMS) and 0.31–0.44 m (maximum) were obtained for the rear cameras, and 0.30-0.59 m (RMS) and 0.36–1.16 m for the side cameras (for all baselines). The rear cameras were confirmed to be the more accurate configuration.

5. Results from other tests were shown to be negatively affected by small ($< 0.2^{\circ}$) misalignments in camera mounting, which gave rise to errors of up to 1.2 m. A correction model was proposed to validate this hypothesis, through which RMS errors were reduced and comparable to the left roundabout tests. A very stiff stereo camera mount with sub-0.5 m baseline is recommended for future experiments.

6. Large stereo baselines were found to have a small improving effect on feature matching, but an overall negative effect on errors due to an increased chance of camera misalignment and difficulty in calibration.

7. Sensitivity to variations in scenery and lighting was observed, which could be addressed with automatic real-time camera adjustment in future work.

7.8 Figures

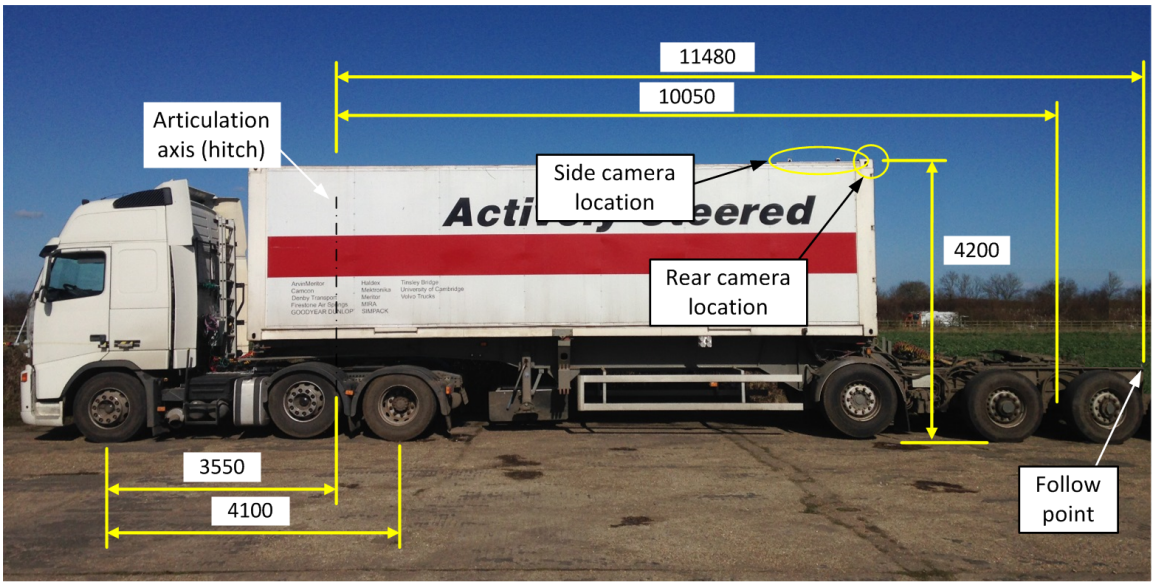


Fig. 7.1 Test vehicle combination, showing camera mounting locations (dimensions in mm)

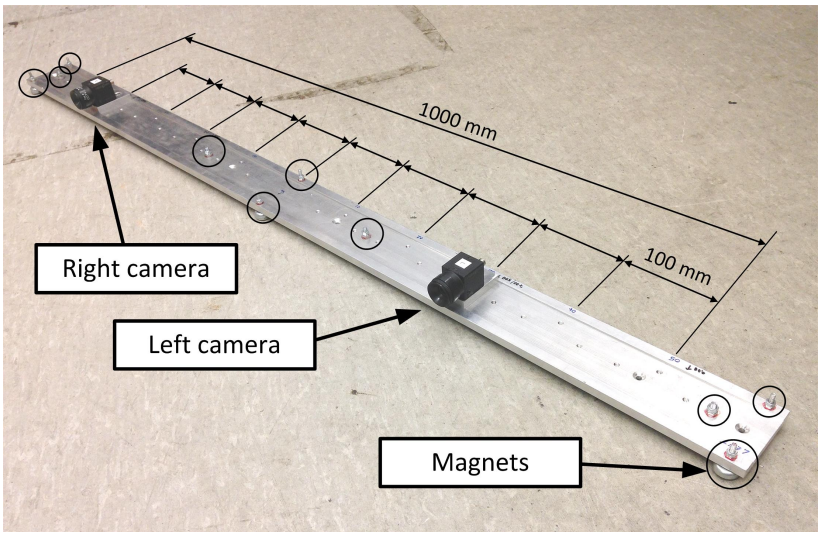


Fig. 7.2 Stereo camera mount

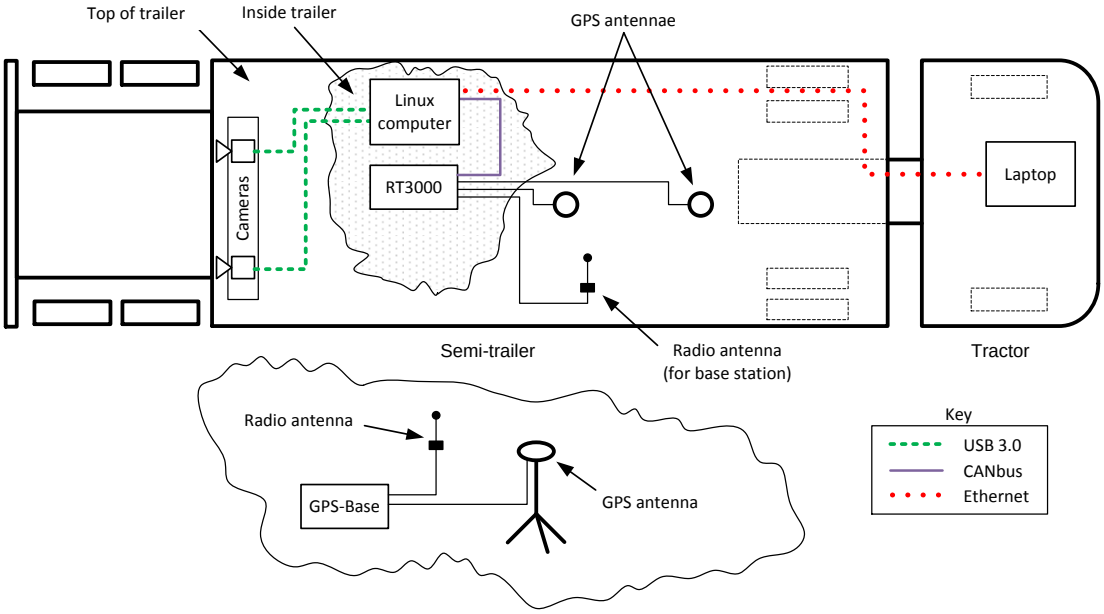


Fig. 7.3 Vehicle instrumentation layout

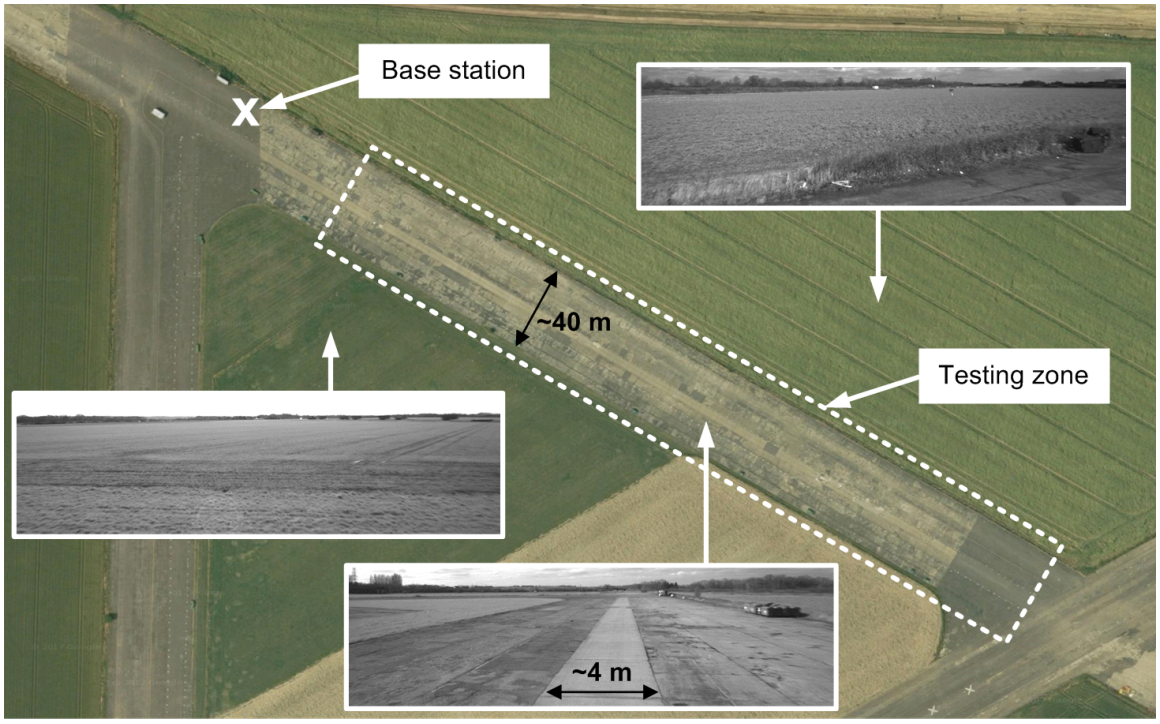


Fig. 7.4 Test site at Bourn airfield, showing sample images of the scenery.

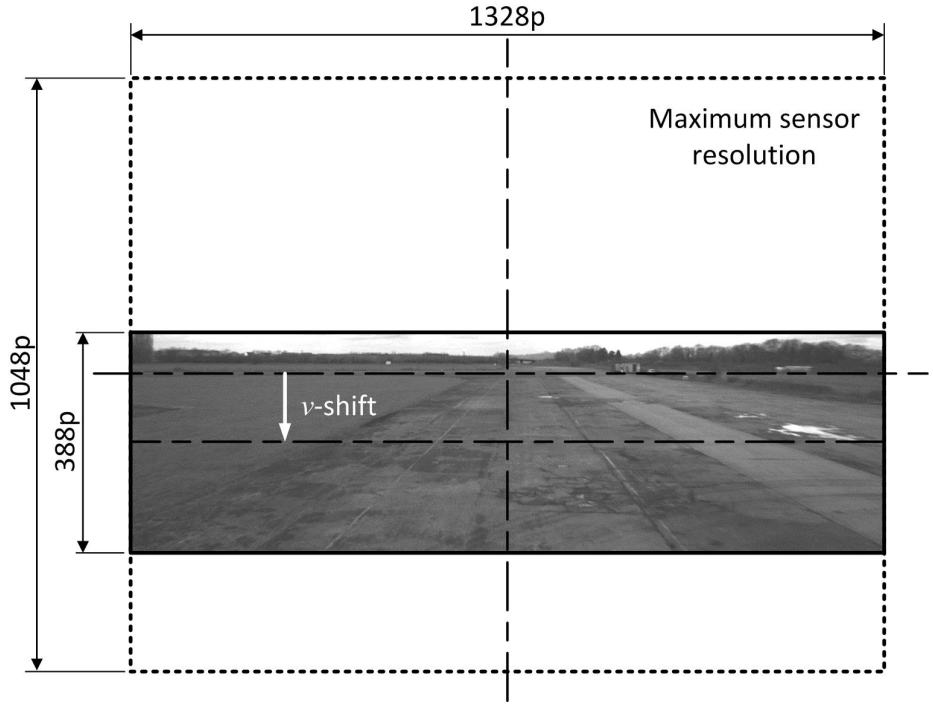


Fig. 7.5 Image resolution and region of interest

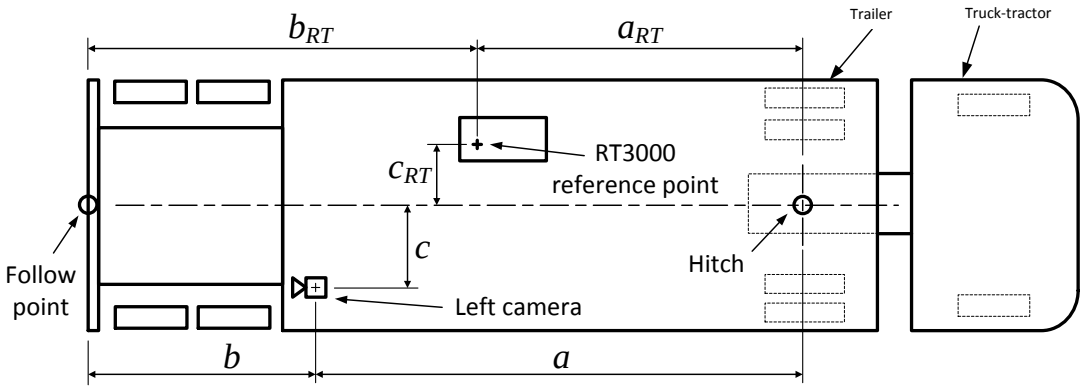


Fig. 7.6 Definition of variables used to denote mounting locations of RT3022 and cameras

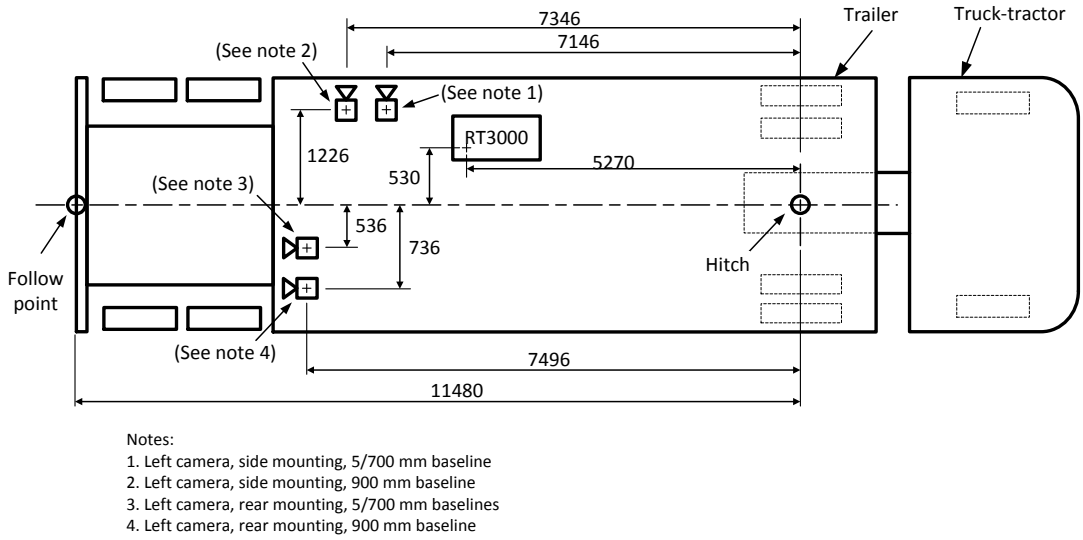
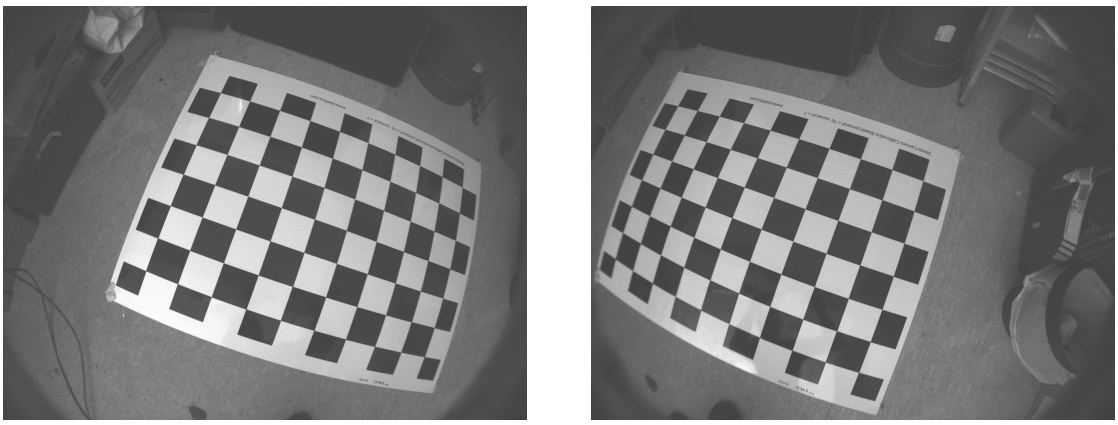
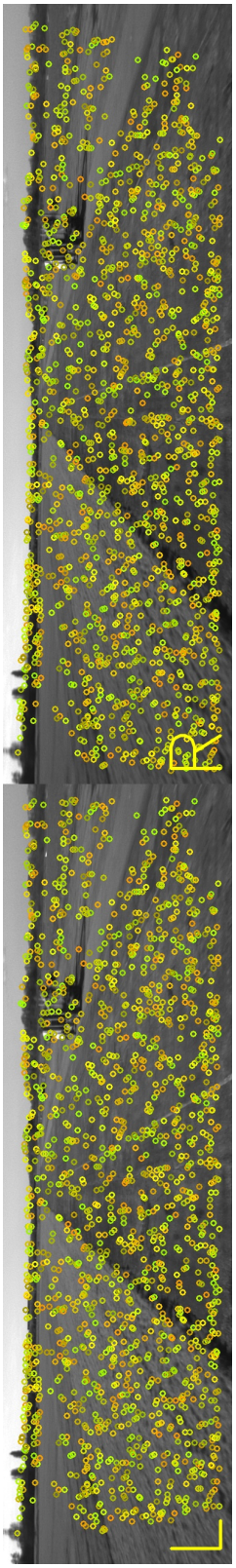


Fig. 7.7 Trailer top view, showing mounting locations of the RT3022 and cameras (dimensions in mm, left camera shown only).

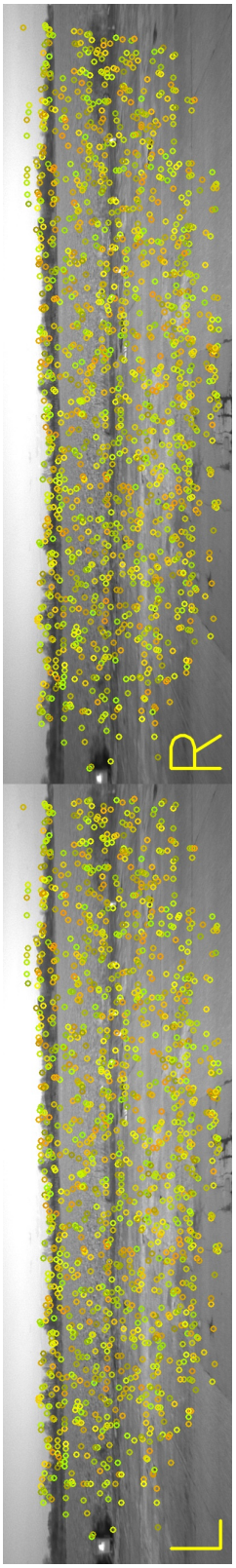


(a) Left camera (b) Right camera

Fig. 7.8 Example stereo calibration image pair, 500 mm baseline

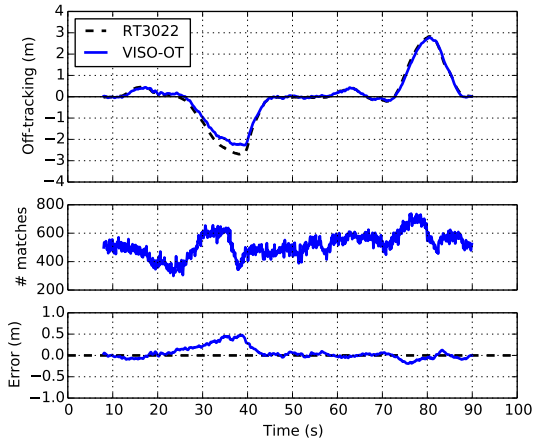


(a) Rear cameras, 900 mm baseline

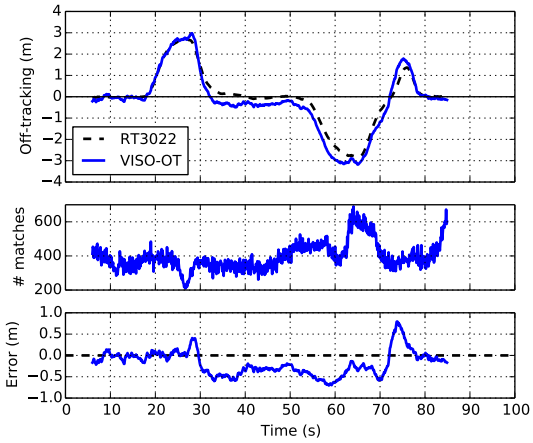


(b) Side cameras, 900 mm baseline

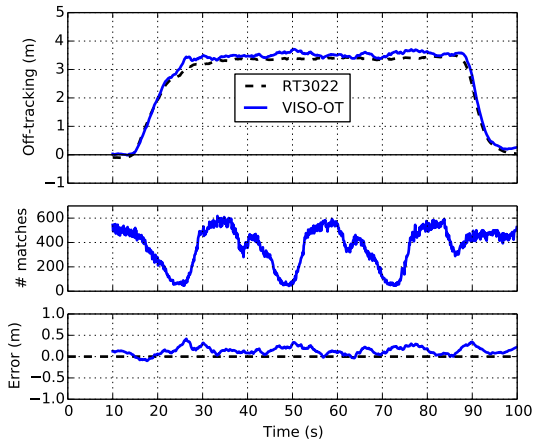
Fig. 7.9 Stereo image pairs with overlaid feature matches, shown at the start of the figure-of-eight manoeuvre



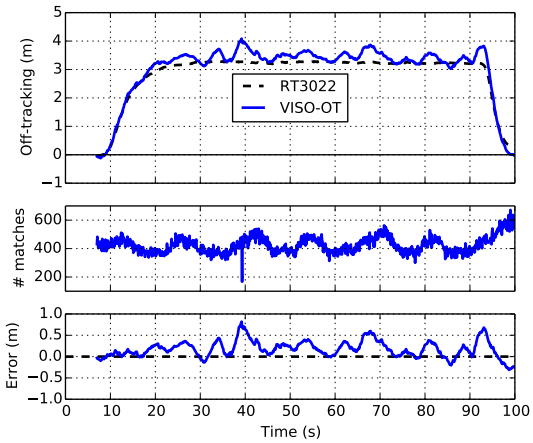
(a) Rear cameras, figure-of-eight



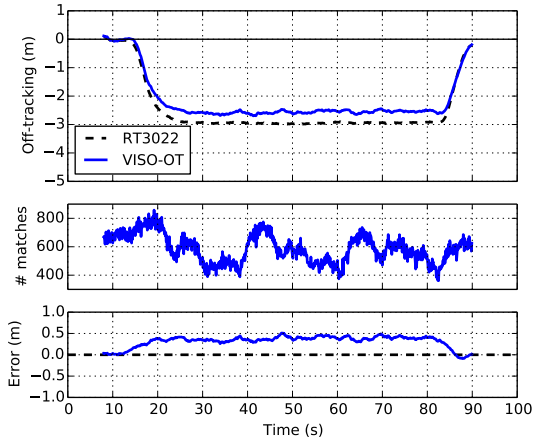
(b) Side cameras, figure-of-eight



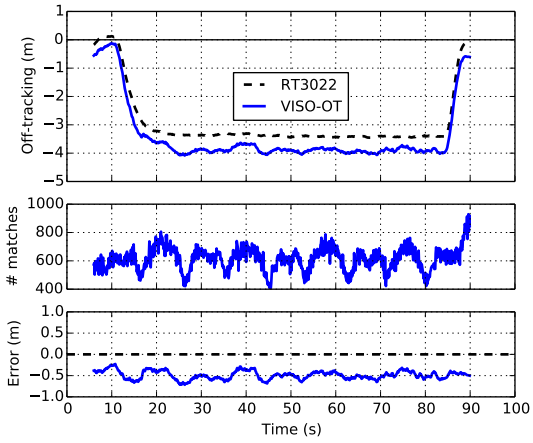
(c) Rear cameras, left roundabout



(d) Side cameras, left roundabout

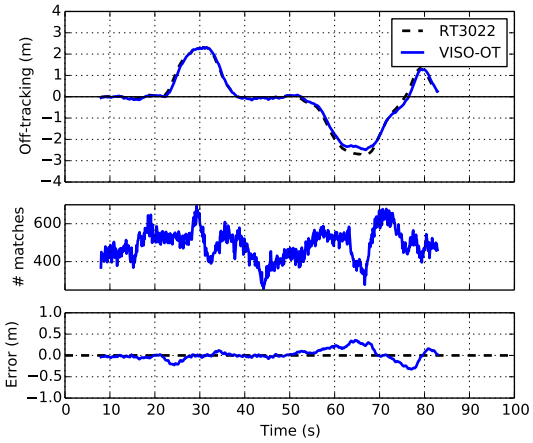


(e) Rear cameras, right roundabout

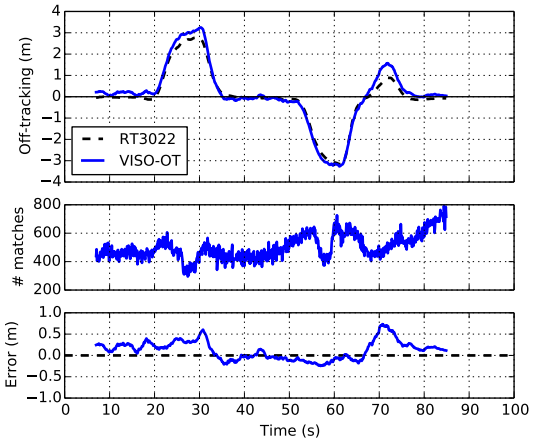


(f) Side cameras, right roundabout

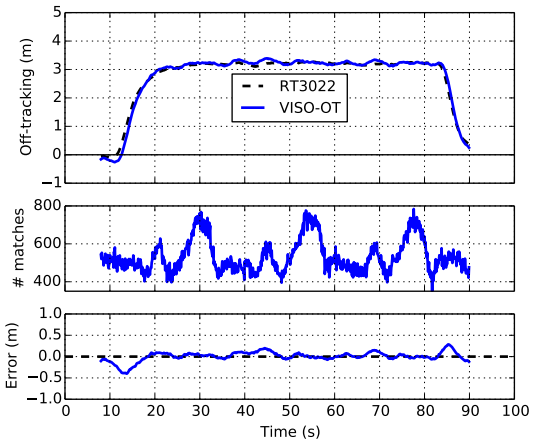
Fig. 7.10 Off-tracking time histories with features and errors, 500mm stereo baseline



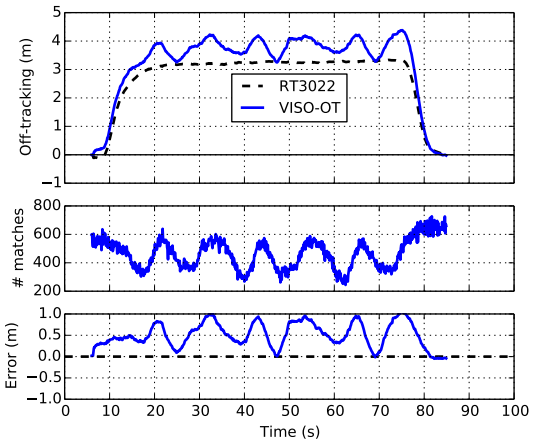
(a) Rear cameras, figure-of-eight



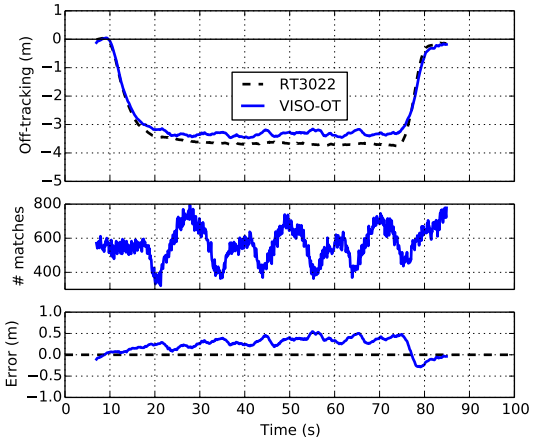
(b) Side cameras, figure-of-eight



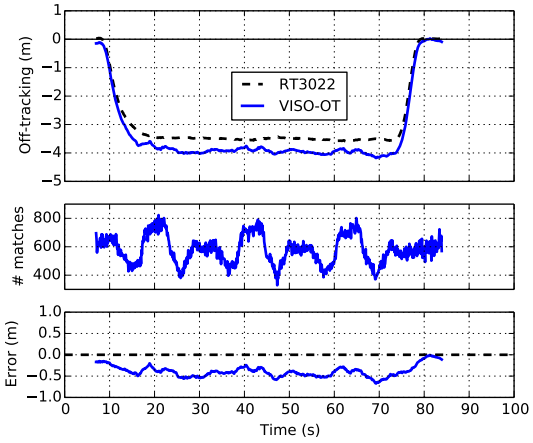
(c) Rear cameras, left roundabout



(d) Side cameras, left roundabout

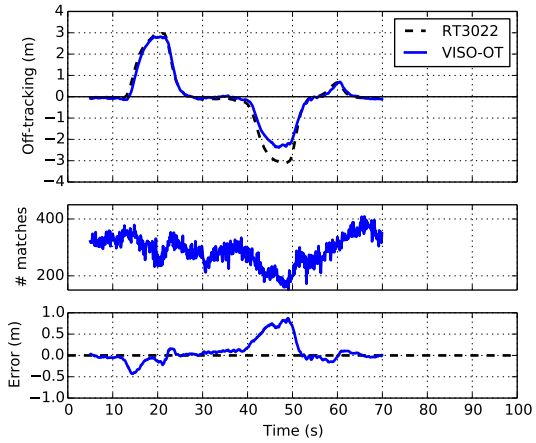


(e) Rear cameras, right roundabout

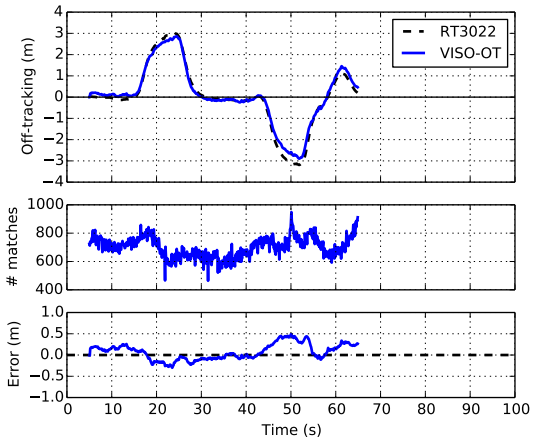


(f) Side cameras, right roundabout

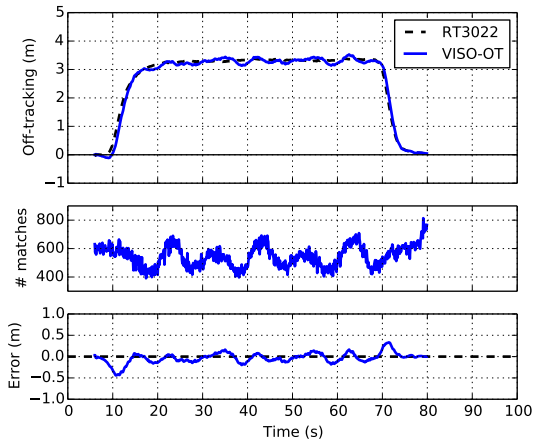
Fig. 7.11 Off-tracking time histories with features and errors, 700mm stereo baseline



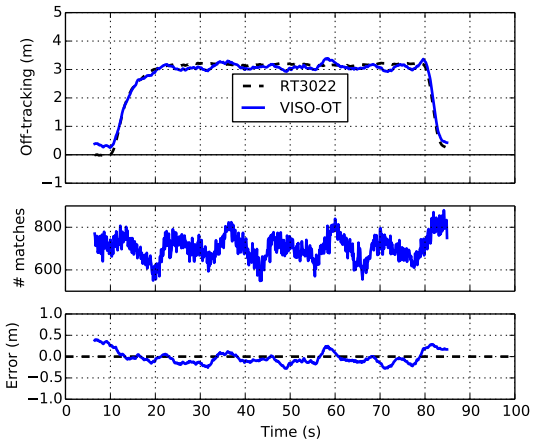
(a) Rear cameras, figure-of-eight



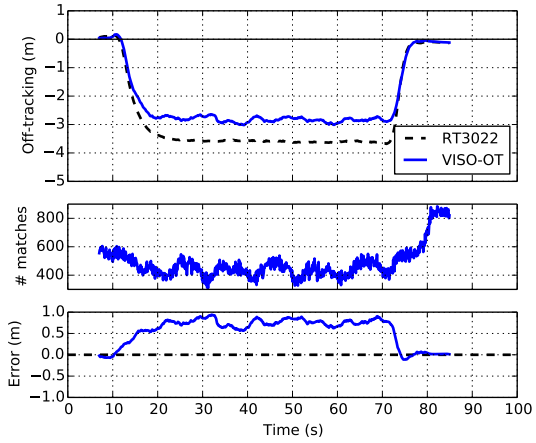
(b) Side cameras, figure-of-eight



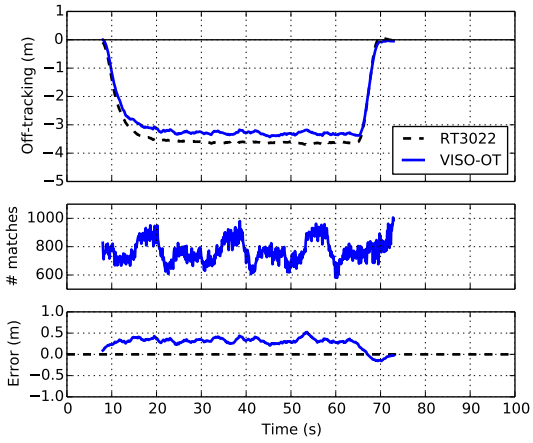
(c) Rear cameras, left roundabout



(d) Side cameras, left roundabout

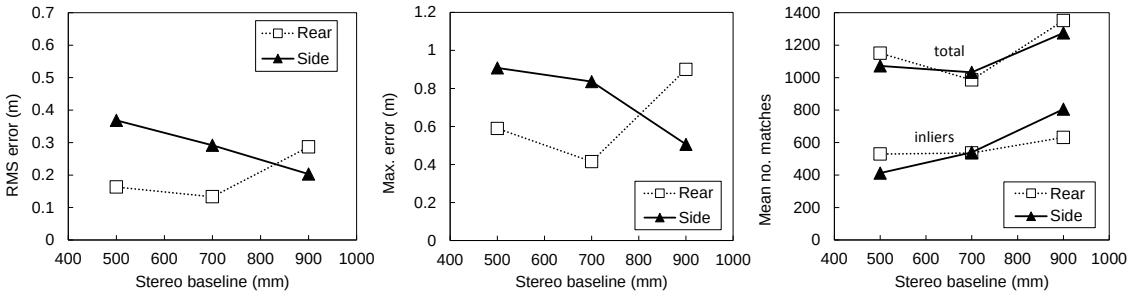


(e) Rear cameras, right roundabout

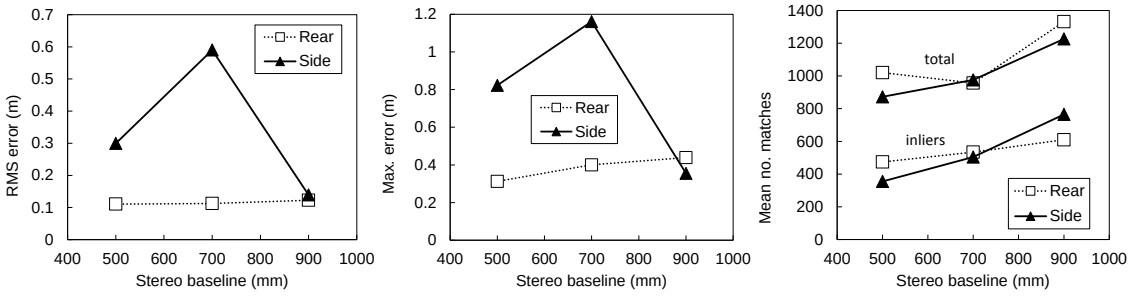


(f) Side cameras, right roundabout

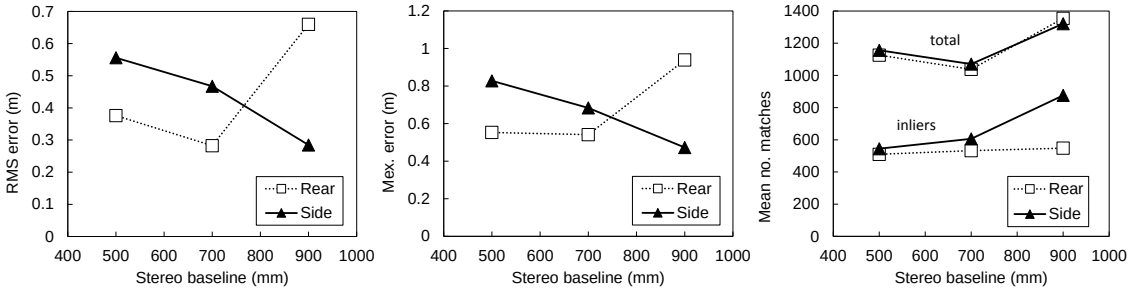
Fig. 7.12 Off-tracking time histories with features and errors, 900mm stereo baseline



(a) Figure-of-eight

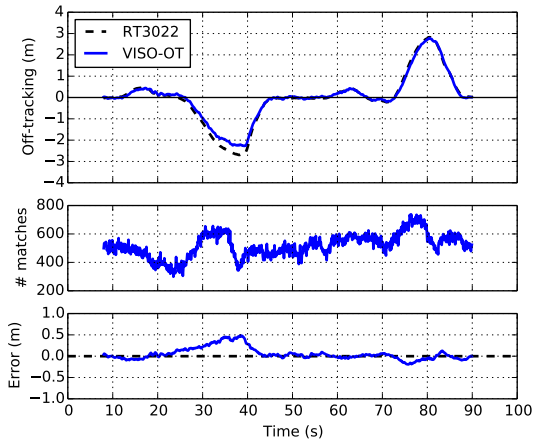


(b) Left roundabout

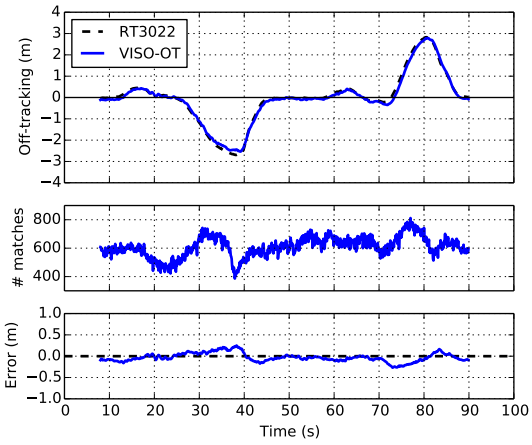


(c) Right roundabout

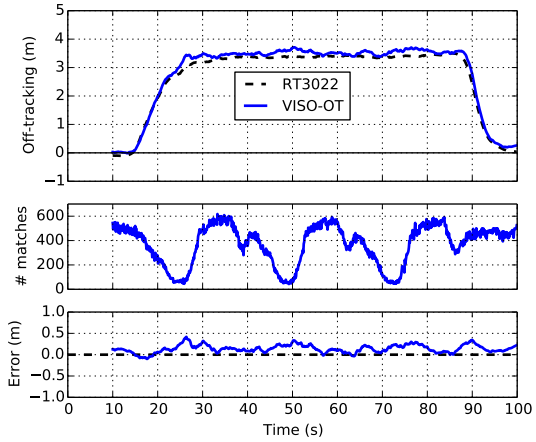
Fig. 7.13 Off-tracking vehicle test results summary



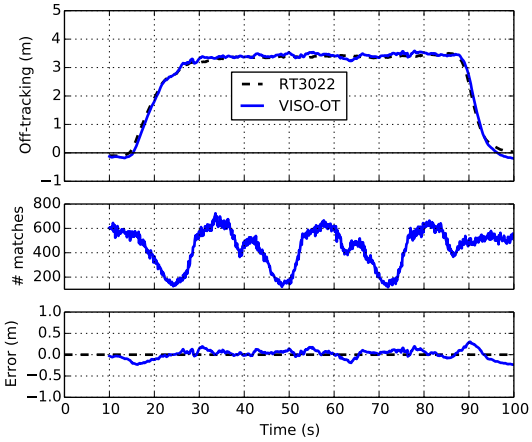
(a) Figure-of-eight



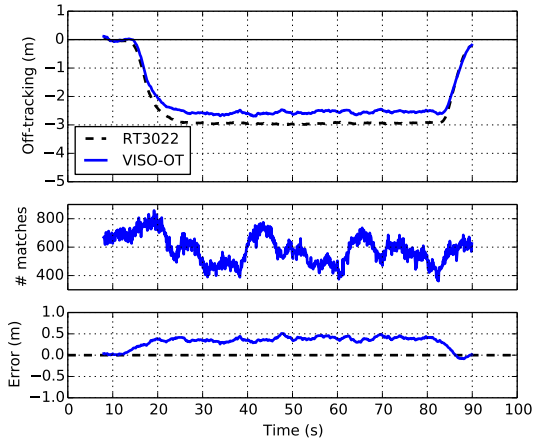
(b) Figure-of-eight, corrected



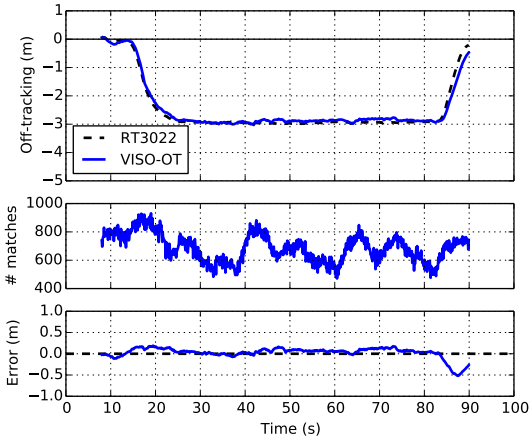
(c) Left roundabout



(d) Left roundabout, corrected

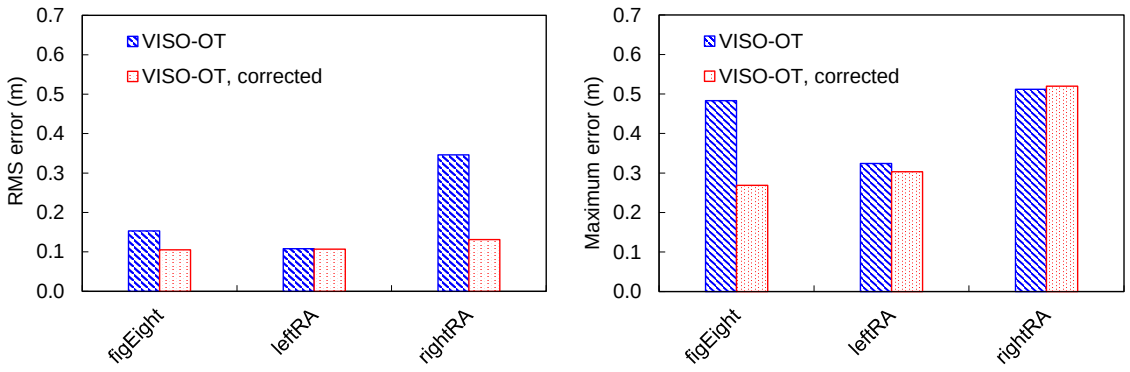


(e) Right roundabout



(f) Right roundabout, corrected

Fig. 7.14 Off-tracking vehicle test results, rear cameras, 500 mm baseline, showing the effect of a simulated rotation of the right camera ($\phi = 0.15°$, $\psi = -0.10°$, $\theta = 0°$)



(a) RMS errors

(b) Maximum errors

Fig. 7.15 The simulated effect of camera rotation ($\phi = 0.15°$, $\psi = -0.10°$, $\theta = 0°$) on errors, rear cameras, 500 mm baseline

Chapter 8

Conclusions and future work

In this chapter, the main conclusions from each chapter are summarised, followed by a review of the research objectives. Lastly, a detailed discussion of recommendations for future work is provided.

8.1 Summary of main conclusions

8.1.1 Introduction and literature review (Ch. 1)

The use of High Capacity Vehicles (HCVs) can yield significant emission reductions in the freight transport sector, with relatively few barriers to introduction compared with alternatives such as improved engine technologies and alternative fuels. However, enabling greater uptake of multiply-articulated HCVs, on more of the road network, requires addressing the manoeuvrability and stability limitations associated with these vehicles.

Active control systems including autonomous reversing, anti-jackknife control, combined braking and steering control, and path-following trailer steering, are important enabling technologies for HCVs which help to address these limitations. Current articulation angle sensing and trailer off-tracking estimation methods present particular barriers to the adoption of these technologies, and hence of HCVs.

A need was identified to develop a non-contact, tractor-based articulation angle sensor compatible with multiple trailer combinations. Vision-based technology utilising SLAM (Simultaneous Localisation and Mapping) or template-based methods were shown to be particularly relevant for this task. In addition, the need for a trailer-based off-tracking sensor which is independent of wheel slip conditions was identified, and visual odometry methods were shown to be appropriate for this task.

8.1.2 Development of an articulation angle sensor (Ch. 3)

A camera-based sensing concept for articulation angle measurement was described, using a rear-facing camera mounted behind the truck/tractor. Two image processing methods were investigated for calculating articulation angle from resultant image sequences. The concept was evaluated by simulation using a CAD model of a tractor semi-trailer and a 3-DOF articulated vehicle model.

The first method ('template-matching') was developed based on limited previous work by Harris [73]. The method was refined and an Unscented Kalman Filter was incorporated, resulting in the combined 'TM+UKF' algorithm. The method yielded RMS and maximum errors of 0.30° and 0.73° respectively in simulations for articulation angles up to 50°. However, it is limited in its assumption of a planar trailer face and requires some knowledge of the trailer geometry.

The second method was based upon the Parallel Tracking and Mapping (PTAM) algorithm of Klein and Murray [123], which was modified for the current application of articulation angle measurement. The method yielded RMS and maximum errors of 0.60° and 1.14° in the same simulations.

Although the PTAM-based method was less accurate than the template-matching method in these simulations, it was shown to be significantly more versatile in that it made no assumption of trailer shape or geometry, and was functional at high articulation angles when the front of the trailer is no longer visible. Additional simulations at up to 90° were performed with the PTAM algorithm alone, yielding RMS and maximum errors of 0.79° and 1.56° respectively.

Both methods were shown to give good performance compared to the published state-of-the-art [30, 52, 73] (comparing simulation results only). Compared to the target errors of 0.6° (maximum) and 0.4° (RMS), the TM+UKF errors were close, but PTAM errors were slightly higher.

8.1.3 Field tests of the articulation angle sensor on a tractor semi-trailer (Ch. 4)

The two methods were evaluated in full-scale vehicle tests on a tractor semi-trailer combination. A digital camera was mounted to the tractor cab, capturing greyscale images at a 640×480 resolution. A commercial articulation angle sensor by V.S.E. B.V. [66] was fitted to the semi-trailer for ground truth measurements.

Low-speed manoeuvring tests were conducted at Bourn Airfield near Cambridge. A set of manoeuvres were carried out with articulation angles up to 37° for a planar trailer

front (where the front surface of the trailer facing the camera was mostly flat/planar, with no significant protrusions or non-planar surfaces). Both template-matching and PTAM methods were evaluated for this dataset. RMS and maximum errors of 1.06° and 3.45° were obtained for the template-matching method respectively, and 0.73° and 1.82° for the PTAM method. The template-matching method was shown to have a sensitivity to trailer pitch which was corrected in the results. It also exhibited large errors at moderate articulation angles, shown to be an inherent result of the normalised cross correlation matching process.

Additional tests with a non-planar addition to the trailer front were carried out for further evaluation of the PTAM method, for articulation angles up to 55°. The algorithm achieved RMS and maximum errors of 1.19° and 2.73° respectively.

The performance of both algorithms in these tests was shown to be superior to that of comparable vehicles tests of other methods in the published literature [70, 72, 73] (comparing vehicle test results only). Target errors were again exceeded, however further refinements to the PTAM algorithm could achieve the desired measurement accuracy.

8.1.4 Field tests of the articulation angle sensor on a truck and full-trailer (Ch. 5)

The articulation angle concept was further developed and evaluated for a rigid truck and full-trailer combination. This combination is very common in Scandinavian countries, and comprises two articulation angles to be measured at the drawbar and trailer respectively. Modifications to the articulation angle sensing concept were made to measure these two articulation angles using the single truck-mounted camera.

The PTAM-based method was carried forwards for this task owing to its superior versatility and accuracy in vehicle tests compared to the template-matching method. Regions of interest in the images were used to isolate the drawbar and trailer measurements, and additional initialisation steps were introduced to improve accuracy.

Full-scale vehicle tests were carried out in Hällered, Sweden, on a truck and full-trailer combination in collaboration with Volvo Group Trucks Technology. The camera was mounted relatively low to the rear of the truck chassis, giving a field of view which included both drawbar and trailer. 'Ground truth' measurements were obtained from a rotary potentiometer at the drawbar and from a VSE articulation sensor on the trailer.

Low to moderate speed manoeuvres were carried out, with articulation angles up to 68°. RMS errors of 0.8–1.8° for the drawbar and 1.1–2.4° for the trailer were observed, with maximum errors in the ranges 2.5–4.6° and 3.4–5.1° respectively. These exceed the target errors by some margin, and are mostly the result of increasing errors with articulation angle

and a possible pitch sensitivity. Performance could be improved in future work by addressing these issues.

8.1.5 Development of a trailer off-tracking sensor (Ch. 6)

A trailer-based concept for estimating trailer off-tracking was demonstrated in Chapter 6, utilising a stereo camera pair and the visual odometry algorithm of Geiger *et al.* [133] and denoted 'VISO-OT'. The concept requires no measurements or parameters from the truck or tractor and is independent of wheel slip conditions. The concept was evaluated in simulations utilising a CAD-based virtual environment representative of off-highway scenarios such as rural farm roads.

Initial simulations were conducted to evaluate the suitability of the simulation environment and methodology, which were shown to be acceptable. Simulations were carried out to evaluate the performance of side- and rear-mounted camera configurations and of selected stereo baseline values. Additional simulations were conducted for a variety of operating conditions including various levels of trailer off-tracking and variations in visual scenery.

The concept demonstrated RMS errors of 0.01–0.04 m under realistic operating conditions and stereo baselines. Rear-mounted cameras were shown to provide better accuracy and more consistent performance than side-mounted cameras. Good robustness to scenery detail was demonstrated, and the system performed well in very low texture scenery.

Although it is theoretically possible to improve accuracy with a larger baseline, the effects of this were not observed in the simulation results. In fact a general increase in errors was observed with increasing stereo baseline due to a reduction in feature matching fidelity between left and right cameras.

Temporary drift was evident in some measurements, owing to the nature of the off-tracking calculation, which stores data in a buffer equivalent to the length of the trailer. However, unlike typical visual odometry applications, the drift here was shown to be bounded due to the relative nature of the off-tracking calculations.

8.1.6 Field tests of the trailer off-tracking sensor on a tractor semi-trailer (Ch. 7)

The performance of the off-tracking estimation concept was investigated in full-scale tests on a tractor semi-trailer combination. Two cameras were mounted to an adjustable stereo camera bracket, which was mounted to the roof of the semi-trailer. The mount was magnetic enabling it to be moved between side-facing and rear-facing configurations as necessary.

An RT3022 inertial and GPS navigation unit was fitted to the semi-trailer for 'ground truth' measurements. Dual GPS antennae were used to obtain accurate heading measurements and a ground-based GPS base station was used to improve GPS position measurement accuracy.

Low- to moderate-speed manoeuvres were carried out at Bourn Airfield on dry tarmac. The scenery detail was relatively sparse, providing a challenging environment for the visual odometry algorithm. In left roundabout manoeuvres, errors of 0.11–0.12 m (RMS) and 0.31–0.44 m (maximum) were obtained for the rear cameras, which is close to the target accuracy. Side cameras exhibited errors of 0.30-0.59 m (RMS) and 0.36–1.16 m (maximum). These larger errors were the result of more erratic measurements, and sensitivity to lighting conditions.

Other tests exhibited errors as high as 1 m for both rear and side cameras. It was demonstrated through the use of a correction model that these errors were likely the result of camera misalignment due to mounting inaccuracies. Sensitivity to camera alignment was shown to be severe, with large errors resulting from $< 0.2^{\circ}$ changes in camera orientation. Large baselines were found to have little benefit, but increased the possibility for misalignment. A stiff stereo camera mount with sub-0.5 m baseline is recommended for future experiments.

8.2 Review of objectives

The first objective set out to develop a tractor-based articulation angle sensor. This was achieved, using a camera-based sensor and two alternative image processing algorithms as described in Chapter 3. The use of vision-based technology enabled remote, non-contact sensing of the trailer from the truck or tractor, without the need for any non-standard connections or modifications to the trailer.

Objectives 1(a) and 1(b) required that the sensor should require minimal information from and modifications to the trailer. Both template-matching (TM+UKF) and PTAM-based methods require no modifications to the trailer, either physical or visual. Both algorithms were shown to perform adequately under conditions of minimal trailer texture or markings (Chapter 4). The PTAM-based method exhibited no loss of performance with a near-featureless trailer (Chapter 5). The PTAM algorithm requires no knowledge of trailer parameters or behaviour, while the template-matching method assumes basic geometric knowledge of the trailer, including hitch location and front overhang.

Objective 1(c) required the system to be compatible with multiple truck and trailer combinations. The concept was demonstrated for a tractor semi-trailer combination in Chapters 3 and 4, the most common type of articulated HGV. In Chapter 5 the system was demonstrated for a 'Nordic combination' (comparable to a truck and full-trailer combination).

By extension, the system is also compatible with truck and drawbar trailer combinations. Compatibility with multi-trailer B-combinations is not yet feasible using a single camera mounted on the tractor, due to the obscured view of the second trailer presented by the leading B-link trailer. The concept may be developed further to accommodate such combinations.

Objective 1(d) pertained to the accuracy of the articulation angle sensor, to provide measurements at 10 Hz with RMS and maximum errors not exceeding 0.4° and 0.6° respectively. The template-matching algorithm achieved a measurement rate of 6–10 Hz, slightly off the target, while the PTAM-based algorithm achieved 20 Hz which is well within the requirement. In terms of measurement accuracy, PTAM was shown to perform best in vehicle tests compared to template-matching, consistently providing the lowest errors. In tractor semi-trailer tests, PTAM yielded errors of 0.4–1.6° (RMS) and 1.2–3.8° (maximum). In tests with a Nordic combination, errors were 0.8–1.8° (RMS) and 2.5–4.6° (maximum) for the drawbar, and 1.1–2.4° (RMS) and 3.4–5.1° (maximum) for the semi-trailer. These achieved measurement accuracies are close to the targets, even meeting them in certain tests, but future work is needed to properly understand the source of these errors and hence improve accuracy.

Objective 2 was to evaluate the articulation angle sensor concept in full-scale vehicle tests. Full-scale tests were conducted on a tractor semi-trailer in Chapter 4 and on a full-scale truck and full-trailer combination in Chapter 5, with performance as presented above.

Objective 3 set out to develop a trailer-based, trailer off-tracking estimation concept. This was achieved, using a stereo camera sensor and the visual odometry algorithm of Geiger *et al.* [133]. Concept development is outlined in Chapter 6.

Objective 3(a) required the off-tracking estimation concept to require minimal information and measurements from the truck or tractor. The proposed concept meets this objective, addressing important limitations of Cheng's estimation technique [32] and Miao's SC-GWNS [3]. The current system only requires trailer knowledge regarding trailer length and the camera mounting location on the trailer.

Objective 3(b) required the system to make no assumption regarding the state of wheel slip, thereby enabling it's use in off-highway and low friction environments. This objective has been achieved through the use of non-contact vision-based sensing, and by avoiding the need for a full vehicle motion model.

Objective 3(c) pertained to the accuracy of the off-tracking sensor, to provide off-tracking measurements at 10 Hz with RMS and maximum errors of less than 0.10 m and 0.15 m. The proposed system is able to provide measurements at 10 Hz, though the target accuracy has not yet been achieved. RMS and maximum errors of 0.11–0.12 m and 0.31–0.44 m were obtained in limited tests on a tractor semi-trailer combination which is near to the target.

However, considerably larger errors were observed in other tests. This was shown to be the result of a pronounced sensitivity to camera alignment.

Objective 4 was to evaluate performance of the system in full-scale vehicle tests. This was achieved, and tests on a full-scale tractor semi-trailer combination were presented in Chapter 7, with performance as presented above.

8.3 Recommendations for further work

8.3.1 Robustness to lighting and weather conditions

One of the main deficiencies of vision-based sensing is its sensitivity to lighting and weather conditions, such as operating under low light conditions at night, during adverse weather such as rain or fog, or under direct sunlight.

Low light conditions mean that less light is able to reach the image sensor. This can be compensated by automatically reducing the camera shutter speed and/or increasing the 'ISO' level (sensitivity of the camera sensor). However, the extent to which shutter speed can be reduced is limited by the desired processing speed and possible motion blur.

For example, if measurements at 10 Hz are required, allowing 100 ms per frame, and the image processing pipeline requires 80 ms of this, then shutter time would need to remain below 20 ms. Additionally, motion blur may become evident at slower shutter speeds, presenting another upper limit on shutter time. This will depend on the anticipated speed of translation of the image features, and on other camera exposure settings.

Increasing the ISO level of a camera sensor allows more light to be captured for a constant shutter speed, but at the expense of increased noise in the image. A given camera sensor also possesses a maximum ISO limit. High ISO low-light cameras are commercially available, but at increased cost.

Image processing algorithms exist which are able to compensate for low lighting conditions to some degree. The extent of this is limited however, and the additional processing time may limit the achievable frame rate.

The addition of artificial light is a potential solution, and has been used in many applications. For the articulation angle sensor, where the target trailer is within a few meters of the camera, adding a few LEDs near the camera source may be sufficient. However, for the off-tracking cameras where the observed environment ranges from 10 m to infinity, this solution is not viable. It is also important to keep in mind country-specific legislation which may prevent the addition of light sources other than for purposes designated in the relevant regulations.

It is suggested that tests be carried out to determine the minimum levels of light under which the sensors would be expected to function, *i.e.* a rural road at night with no road lighting or other traffic. The only source of light would be ambient light generated and reflected by the vehicle itself, such as from the truck headlamps and from the truck and trailer tail lights.

For the articulation camera, these ambient lighting conditions would differ for a tractor semi-trailer combination, where the tail lights of the tractor would be beneath the trailer and hence obscured, and a rigid truck and trailer combination, where the truck tail lights would provide some illumination to the drawbar and trailer.

For the off-tracking cameras, ambient illumination from the vehicle would be significantly less. The greatest source of illumination would likely be light from the truck headlamps and truck/trailer tail lights, reflected from the nearby roadside scenery. For this reason it may be beneficial for the cameras to be side-facing, and mounted towards the front of the trailer where the reflected light would be greatest from the headlamps, or rear-facing and mounted at the rear of the trailer where light from the tail lights would be greatest. Additionally, introducing a downwards tilt to the cameras may prove beneficial to utilise the nearest most illuminated scenery or road surface for feature detection. Alternatively, the cameras could be mounted lower down to the rear of the trailer chassis, viewing low to medium distance ground features illuminated artificially.

Once the minimum levels of anticipated lighting have been identified, appropriate technologies should be investigated to meet these requirements. For the articulation camera this can likely be compensated using a few LEDs next to the camera. For the off-tracking cameras it is likely that cameras with greater low-light sensitivity may be required.

An alternative option to consider is active infrared camera sensing. An infrared camera (or a conventional camera with an infrared filter) together with an infrared emitter next to the camera could be used for low light conditions. The emitter and filter could be disabled during good lighting, or the exposure of the camera could be automatically compensated to function during daylight.

Lighting conditions such as rain or fog may obscure the scenery to some extent, impacting feature detection, however the effects of this are expected to be minimal except in extreme cases. In any case, if vision is severely limited due to extreme weather, safety would dictate that vehicles significantly reduce speed or stop until conditions improve.

The accumulation of water on the camera lens may have more of an impact, blurring and distorting the images. However, numerous existing solutions to this problem could be utilised. Volvo Truck's reverse camera [193] features a protective electronic visor and a

heated camera to prevent fogging. Self-cleaning camera technology is also commercially available, from Orlaco for example [194].

The prospect of dirt and mud collection on the camera lens could be similarly addressed. However, the potential for mud splash on the articulation or off-tracking cameras is small compared with the ground-watching cameras proposed by Miao [3]. A self-cleaning camera such as [194] would address this issue as well should dirt accumulate over time.

Finally, the image processing techniques adopted (PTAM and visual odometry) are not necessarily restricted to vision-based inputs. As was described in Section 1.7.1, active technologies such as lidar have also been used to provide sensor inputs to such algorithms. Although lidar was dismissed for this work primarily due to its expense, its cost will reduce further as the technology is developed further, and so it may become relevant to explore its use for these tasks.

8.3.2 Closed-loop testing of the articulation angle sensor

The intended applications of the articulation angle sensor are for autonomous reversing, combined braking and steering, and anti-jackknife control. Closed-loop tests (tests in which the output of the measurement is used as the controller input) with one or more of these technologies should therefore be explored.

In high-speed applications such as combined braking and steering, closed-loop control would ensure that articulation angles remained within a much narrower range than has been evaluated in this work. The accuracy of the sensor was shown to be best at smaller angles, and so closed-loop performance may be better than open-loop performance suggests. However, for autonomous reversing technology such as [47], large articulation angles are expected. Anti-jackknife control does not require high accuracy, as it is only concerned with the onset of a threshold articulation angle.

8.3.3 Articulation sensing for B-trailer combinations

The articulation angle sensor has been shown to be functional for a number of HGV combinations including tractor semi-trailer, truck and full-trailer, Nordic, and truck and drawbar combinations. However, it is limited in its application to B-trailer combinations with more than one trailer due to the first trailer obscuring the second and subsequent trailers. The B-double combination is very popular in a number of countries, and so it is important that feasibility with this and similar combinations is explored.

The versatility of the PTAM algorithm makes it attractive for this task, and it could theoretically be suitable with relatively minor modifications. However, the primary challenge

is a practical one, in finding a suitable mounting location such that all trailers are in the view of the articulation measurement camera.

Mounting options for this task may include a raised camera, viewing the full length of trailers from an elevated viewpoint on the tractor. In most applications however, this may not prove practical given that trailers are typically as tall as or taller than truck and tractor cabs. Furthermore, most countries regulate total vehicle height and often trailers are built to this limit, presenting further limitations. The UK is an exception however, with no legislated limit on HGV height [18].

Another option is side-mirror-mounted cameras, able to infer articulation angle from an oblique side view of the trailers. One camera per mirror would be necessary to account for the full range of trailer motion. This is similar to the configuration proposed by Lu *et al.* [68], however their patent assumes the knowledge of trailer wheelbase and camera location, whereas PTAM could perform this task without this knowledge. Vertically-separated regions of interest would be required to distinguish individual trailers, however these regions would move with trailer motion and so additional work would be needed to enable this functionality.

8.3.4 Other applications for the articulation angle sensor

A rearward-facing camera mounted to a truck or tractor can serve multiple uses, some of which are already commercially available. 'Exterior vision cameras' and 'reversing cameras' are available on a number of passenger vehicles and on some trucks [193]. These are often analogue cameras and only provide field of view support to the driver during reversing, parking, or trailer hitching.

In a similar manner, the live video feed of the articulation angle sensor could be relayed to the driver whilst also performing articulation measurements as necessary. Relative articulation and translation measurements of an uncoupled trailer could also be used to provide quantitative visual support to the driver to aid trailer coupling. Modern passenger vehicles have a similar feature in which a reversing camera indicates future vehicle trajectory based on current steer angles.

The camera could also form part of a wider camera array providing a 'bird's-eye view' functionality. This helps alleviate blind-spots, and can serve as an additional reversing and manoeuvring aid to the driver. Commercial systems to this effect are available from Continental [195], Autowatch [196] and Safety Shield Systems [197]. Related research includes the work by Ehlgen [70] and Chang [198].

Reversing controllers such as that of Rimmer *et al.* [29] may require knowledge of vehicle parameters other than articulation angle depending on the controller employed. In

these cases it may be beneficial to obtain additional trailer information from the articulation camera, such as hitch location and trailer length.

The PTAM-based articulation camera could be used for these purposes with some modification, depending on the camera-mounting configuration adopted. Hitch location could be measured with the current proposed mounting location of the camera, but length estimation may require elevated or side-mirror mounting options. The current mono camera implementation of the concept is unable to infer distance and translation without external information; this could be remedied by assuming a standardised trailer to scale translation measurements. Alternatively a stereo-camera implementation such as S-PTAM [125] could be adopted which does not suffer from the same limitation.

Lastly, more generalised absolute rotation measurements could be obtained from a PTAM-like algorithm, *i.e.* measurements which are not restricted to rotations in a given plane. This may prove useful in future work on general motion and pose estimation of trailers as well as in possible related applications.

8.3.5 Closed-loop testing of the off-tracking sensor

Miao developed an off-tracking measurement concept using two ground-watching cameras beneath the trailer [3]. This 'ground-watching navigation system' was assessed in closed-loop tests on dry tarmac. Measurements from the ground-watching system were used as the input to Cheng's [32] path-following trailer steering controller, and implemented on a steered trailer. Miao's modification to Cheng's path-following controller [3, 32] could be used in future work, as this incorporates the use of an independent off-tracking measurement for trailer steering control.

The new visual odometry-based off-tracking estimation concept should be similarly assessed, investigating the overall closed-loop performance of path-following trailer steering when using these external off-tracking measurements. However, where Miao conducted tests on dry tarmac, the proposed investigation would benefit from tests on a road surface representative of challenging off-highway conditions with a combination of low friction, camber and grade. This would promote simultaneous off-tracking and wheel slip. Additionally, the surrounding road-side scenery should be representative of typical off-highway environments to assess the robustness of the visual odometry-based measurements.

Miao's ground-watching system demonstrated good accuracy on dry tarmac, which the visual-odometry method in its current form has not been able to surpass. However, Miao assumed a flat and unchanging road surface which may affect the system's performance on off-highway road surfaces. For this reason it may be beneficial to reassess Miao's system on the suggested representative road surface, so that a direct performance comparison with

the visual odometry-based system may be made. Comparisons with Cheng's original state estimation technique for off-tracking measurement could also be made.

8.3.6 Off-tracking sensing for multiple trailers

Both simulation and vehicle testing results (Chapters 6 and 7) showed that rear-mounted cameras yielded better accuracy and robustness compared to side-mounted cameras. However, it was also highlighted that side cameras present a practical advantage over rear cameras in multi-trailer combinations. It would therefore be valuable to investigate more closely the source of errors of the side camera configuration, so as to achieve adequate accuracy for control application.

Alternatively, a compromise between rear-and side-mounted cameras in which the cameras are mounted at some angle between these extremes may prove to be practical. As is clear from Chapter 6, the more rear-facing the cameras the better, but sufficient angular offset should be incorporated such that subsequent trailers do not appear (significantly) in the field of view of the cameras.

A related option lies in the use of a skewed stereo camera configuration, in which the cameras could be side-mounted, but both rotated individually towards the trailer rear by the same degree (see for example [199]). This would require modifications to the visual odometry algorithm.

It is interesting to bring articulation angle sensing into the discussion at this point, and to revisit the concept of tractor- and trailer-based control technologies. Trailer steering control is a trailer-based technology, and so for this application it may be reasonable to assume that trailer-based articulation angle sensing may be available (such as the VSE sensor). It is a reasonable assumption that all steered trailers could adopt articulation angle sensing as standard (such as in [66]).

If all trailers in a vehicle combination are steered trailers, and all possess articulation angle sensing, then the total off-tracking of the combination can be calculated with only one pair of off-tracking cameras on one of the vehicle units, which could be the tractor or the rearmost trailer. This is similar to Miao's Single Camera Ground Watching Navigation System (SC-GWNS). The total measurement error of such a system would increase with each additional trailer (and hence articulation angle sensor) used, but provided high-precision sensors are used this solution may prove adequate.

8.3.7 Other applications for the off-tracking sensor

The stereo camera pair for off-tracking measurement is a potentially useful tool for a number of other measurement tasks. For reversing applications the stereo cameras could be used for obstacle detection and path-planning. This would of course require additional communication between the trailer and the truck or tractor, but may be an attractive area of development in the future particularly with the availability of wireless communications. Visual odometry measurements from the cameras could also be used to detect trailer-crabbing (and hence axle misalignment for example), or for state measurements for electronic stability control applications.

The uses for a stereo camera pair on the truck or tractor unit are similarly numerous and worth investigation. They could be used for obstacle detection, for path-planning, inertial measurement, to support and refine GPS measurements or to replace GPS in GPS-deprived environments such as mines or logging forests. Mature algorithms for these applications exist in the literature for application to passenger vehicles, but extension and application to HGVs has not been investigated in depth before.

If a truck or tractor was fitted with both *tractor-based* off-tracking and *tractor-based* articulation angle sensing, it would be possible to determine trailer off-tracking. The length of the trailer/s would need to be known however. This is an alternative configuration which may prove attractive to truck manufacturers for certain applications.

8.3.8 Measurement errors

The PTAM-based articulation angle sensor was shown to exhibit increased errors with articulation angle, which was attributed to some combination of pitch angle sensitivity, feature distribution, and inherent traits of the PTAM algorithm. Further investigation into these errors could help identify ways to improve measurement accuracy and achieve the target precision of 0.4° (RMS error) and 0.6° (maximum error).

Similarly, the off-tracking sensor was shown to be sensitive to errors in camera alignment and lighting variations. Future experiments should utilise a very stiff stereo camera mount, with a sub-0.5 m stereo baseline, in order to minimise misalignment. Automatic camera adjustment algorithms should be incorporated to reduce sensitivity to lighting variations. With these changes, it is expected that the target measurement precision of 0.10 m (RMS error) and 0.15 m (maximum error) is achievable.

References

[1] J. Aurell and T. Wadman, "Vehicle combinations based on the modular concept," Nordic Road Association, Tech. Rep., 2007.

[2] B. Jujnovich, R. L. Roebuck, A. Odhams, and D. Cebon, "Implementation of active rear steering of a tractor - semi-trailer," in *Proc. 10th Int. Symp. Heavy Veh. Transp. Technol.*, Paris, pp. 358–367.

[3] Q. Miao, "Vision-based path-following control of articulated vehicles," PhD thesis, University of Cambridge, 2015.

[4] G. Kregoski, "FMCW radar in automotive applications," 2016. [Online]. Available: https://www.cst.com/~/media/cst/landing-pages/2016automotive/understanding--testing-fmcw-automotive-radar-devices.ashx?la=en (Accessed 2017-06-05).

[5] J. Lowensohn, "Behind the wheel: a look inside Google's self-driving cars," 2014. [Online]. Available: http://www.theverge.com/google/2014/5/14/5714602/photos-inside-googles-self-driving-cars (Accessed 2013-10-04).

[6] I. Feige, "Freight transport and economic growth – two inseparable twins?" in *Transp. trade Econ. growth - coupled or decoupled?* Berlin Heidelberg: Springer, 2007, pp. 9–73.

[7] "Domestic road freight statistics, United Kingdom, 2015," Department for Transport, London, Tech. Rep. August, Aug. 2016.

[8] H. Baker, R. Cornwell, E. Koehler, and J. Patterson, "Review of low carbon technologies for Heavy Goods Vehicles - Annex 1," Department for Transport, London, Tech. Rep. June, 2009.

[9] D. Kay and N. Hill, "Opportunities to overcome the barriers to uptake of low emission technologies for each commercial vehicle duty cycle," The Low Carbon Vehicle Partnership, London, Tech. Rep., Nov. 2012.

[10] D. Z. Leach and C. J. Savage, "Impact assessment : High Capacity Vehicles," University of Huddersfield, Huddersfield, Tech. Rep., 2012.

[11] "Smart Truck (PBS) pilot project: Status report," CSIR South Africa, Pretoria, Tech. Rep., Feb. 2017.

[12] "Moving freight with better trucks," International Transport Forum/OECD, Paris, Tech. Rep., Apr. 2011.

[13] "Performance based standards scheme – The standards and vehicle assessment rules," National Transport Commission, Melbourne, Tech. Rep., 2008.

[14] J. J. de Pont, "The development of pro-forma over-dimension vehicle parameters," Auckland, New Zealand, Tech. Rep., 2010.

[15] J. R. Billing and R. B. Madill, "Introduction of long combination vehicles in Ontario," in *11th Int. Symp. Heavy Veh. Transp. Technol.*, Melbourne, Australia, 2010.

[16] P. A. Nordengen, "Monitoring results of two PBS demonstration vehicles in the forestry industry in South Africa," in *11th Int. Symp. Heavy Veh. Transp. Technol.*, Melbourne, Australia, 2010.

[17] B. Kraaijenhagen, T. Barth, K. Kural, J. Pauwelussen, I. Besselink, A. Prati, M. Meijs, and H. Nijmeijer, "Greening and safety assurance of future modular road vehicles: book of requirements," HTAS EMS project consortium, Tech. Rep., 2014.

[18] "The Road Vehicles (Construction and Use) Regulations," pp. 3294–3478, 1986.

[19] A. J. Rimmer, A. Odhams, and D. Cebon, "Autonomous reversing of heavy goods vehicles," in *Int. Symp. Heavy Veh. Transp. Technol.*, Stockholm, 2012.

[20] C. Cheng, A. Odhams, R. L. Roebuck, and D. Cebon, "Feedback control of semi-trailer steering at low speeds," *Submitt. to Proc. Inst. Mech. Eng. Part D J. Automob. Eng.*, 2011.

[21] B. A. Jujnovich and D. Cebon, "Path-Following Steering Control for Articulated Vehicles," *J. Dyn. Syst. Meas. Control*, vol. 135, no. 3, p. 031006, 2013.

[22] R. L. Roebuck, A. Odhams, K. Tagesson, C. Cheng, and D. Cebon, "Implementation of trailer steering control on a multi-unit vehicle at high speeds," *J. Dyn. Syst. Meas. Control*, vol. 136, no. 2, pp. 021 016–021 016–14, Dec. 2013.

[23] A. J. Rimmer and D. Cebon, "Implementation of reversing control on a doubly articulated vehicle," *J. Dyn. Syst. Meas. Control*, vol. 139, no. 6, pp. 061 011–061 011–9, Apr. 2017.

[24] G. Morrison, "Combined emergency braking and turning of articulated heavy vehicles," PhD thesis, University of Cambridge, 2015.

[25] L.-K. Chen and Y.-A. Shieh, "Jackknife prevention for articulated vehicles using model reference adaptive control," *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.*, vol. 225, no. 1, pp. 28–42, Jan. 2011.

[26] "The Logistics Report 2013," Freight Transport Association Limited, Kent, Tech. Rep., 2013.

[27] L. M. Henderson, "Improving emergency braking performance of heavy goods vehicles," PhD thesis, University of Cambridge, 2013.

[28] B. A. Jujnovich, "Active steering of articulated vehicles," PhD thesis, University of Cambridge, 2005.

[29] A. J. Rimmer and D. Cebon, "Theory and practice of reversing control on multiply-articulated vehicles," *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.*, vol. 230, no. 7, pp. 899–913, 2016.

[30] M. Bouteldja, A. Koita, V. Dolcemascolo, and J. C. Cadiou, "Prediction and detection of jackknifing problems for tractor semi-trailer," in *IEEE Veh. Power Propuls. Conf.* Windsor: IEEE, Sep. 2006, pp. 1–6.

[31] G. Morrison and D. Cebon, "Combined emergency braking and turning of articulated heavy vehicles," *Veh. Syst. Dyn.*, vol. 55, no. 5, pp. 725–749, May 2017.

[32] C. Cheng, "Enhancing safety of actively-steered articulated vehicles," PhD thesis, University of Cambridge, 2009.

[33] "A simplified guide to lorry types and weights," Department for Transport, London, Tech. Rep., Oct. 2003.

[34] "Carbon budgets: how we monitor emissions targets," 2016. [Online]. Available: https://www.theccc.org.uk/tackling-climate-change/reducing-carbon-emissions/carbon-budgets-and-targets/ (Accessed 2017-06-06).

[35] J. Gummer, D. Kennedy, S. Fankhauser, B. Hoskins, P. Johnson, J. King, J. Krebs, R. May, and J. Skea, "Meeting carbon budgets - 2013 progress report to parliament," Committee on Climate Change, London, Tech. Rep. June, Jun. 2013.

[36] "Strategy for reducing heavy-duty vehicles' fuel consumption and CO2 emissions," European Commission, Brussels, Tech. Rep., May 2014.

[37] M. Phippard, "B-trains, interlinks and B-doubles." [Online]. Available: http://www.hankstruckpictures.com/martin_phippard_trailers_01.html (Accessed 2017-06-06).

[38] "Smart truck programme: Rules for the development and operation of smart trucks as part of the performance-based standards research," CSIR South Africa, Pretoria, Tech. Rep. November, 2010.

[39] "Council Directive 96/53/EC of 25 July 1996 laying down for certain road vehicles circulating within the community the maximum authorized dimensions in national and international traffic and the maximum authorized weights in international traffic," Brussels, pp. 59–75, 1996.

[40] "European Modular System: News," 2017. [Online]. Available: http://www.modularsystem.eu/en/news/ (Accessed 2017-07-04).

[41] "Trial of Longer Semi-Trailers (LSTs): How to take part in the trial: January 2017," Department for Transport, London, Tech. Rep., Mar. 2017.

[42] P. S. Fancher and A. Mathew, "A vehicle dynamics handbook for single-unit and articulated heavy trucks (UMTRI-87-27)," UMTRI, Ann Arbor, Michigan, Tech. Rep., 1987.

[43] H. Prem, J. J. de Pont, B. Pearson, and J. McLean, "Performance characteristics of the Australian heavy vehicle fleet," National Road Transport Commission, Melbourne, Tech. Rep., 2002.

[44] P. S. Fancher and A. Mathew, "Safety implications of various truck configurations, Vol. 1, Technical report (UMTRI-88-32)," UMTRI, Ann Arbor, Michigan, Tech. Rep., 1988.

[45] C. B. Winkler, P. S. Fancher, and C. C. Macadam, "Parametric analysis of heavy duty truck dynamic stability (UMTRI-83-13), Vol. 1, Technical report," UMTRI, Ann Arbor, Michigan, Tech. Rep., 1983.

[46] J. I. Miller, L. M. Henderson, and D. Cebon, "Designing and testing an advanced pneumatic braking system for heavy vehicles," *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 227, no. 8, pp. 1715–1729, Nov. 2012.

[47] A. J. Rimmer, "Autonomous reversing of multiply-articulated vehicles," PhD thesis, University of Cambridge, 2014.

[48] J. Morales, A. Mandow, J. Martínez, A. Reina, and A. García-Cerezo, "Driver assistance system for passive multi-trailer vehicles with haptic steering limitations on the leading unit," *Sensors*, vol. 13, no. 4, pp. 4485–4498, Apr. 2013.

[49] M. M. Michałek, "A highly scalable path-following controller for N-trailers with off-axle hitching," *Control Eng. Pract.*, vol. 29, pp. 61–73, 2014.

[50] P. Bolzern, R. M. DeSantis, and A. Locatelli, "An input-output linearization approach to the control of an n-body articulated vehicle," *J. Dyn. Syst. Meas. Control*, vol. 123, no. 3, pp. 309–316, 2001.

[51] N. L. Azad, A. Khajepour, and J. McPhee, "Analysis of jackknifing in articulated steer vehicles," in *IEEE Veh. Power Propuls. Conf.* IEEE, 2005, pp. 216–220.

[52] L. Chu, Y. Fang, M. Shang, J. Guo, and F. Zhou, "Estimation of articulation angle for tractor semi-trailer based on state observer," in *Int. Conf. Meas. Technol. Mechatronics Autom.*, vol. 2. Changsha City: IEEE, Mar. 2010, pp. 158–163.

[53] J. Chiu and A. Goswami, "The critical hitch angle for jackknife avoidance during slow backing up of vehicle–trailer systems," *Veh. Syst. Dyn.*, vol. 52, no. 7, pp. 992–1015, Jul. 2014.

[54] F. W. Kienhöfer, "Heavy vehicle wheel slip control," PhD thesis, University of Cambridge, 2011.

[55] F. W. Kienhöfer and D. Cebon, "Improving ABS on heavy vehicles using slip," in *Proc. 9th Int. Symp. Heavy Veh. Weight. Dimens.*, Pennsylvania, USA, 2006.

[56] J. I. Miller, "Advanced braking systems for heavy vehicles," PhD thesis, University of Cambridge, 2010.

[57] L. M. Henderson, J. I. Miller, and D. Cebon, "A high performance slip control braking system for heavy goods vehicles," in *Int. Symp. Adv. Veh. Control*, no. 1. Seoul: The Korean Society of Automotive Engineers, 2012.

[58] L. Henderson and D. Cebon, "Full-scale testing of a novel slip control braking system for heavy vehicles," *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.*, vol. 230, no. 9, pp. 1221–1238, Sep. 2015.

[59] G. Morrison and D. Cebon, "Model-based stabilization of articulated heavy vehicles under simultaneous braking and steering," in *Int. Symp. Adv. Veh. Control*, Tokyo, 2014.

[60] G. Morrison and D. Cebon, "Sideslip estimation for articulated heavy vehicles in low friction conditions," in *IEEE Intell. Veh. Symp.* Seoul: IEEE, Jun. 2015, pp. 65–70.

[61] B. Jujnovich and D. Cebon, "Comparative performance of semi-trailer steering systems," in *Proc. 7th Int. Symp. Heavy Veh. Weight. Dimens.*, Delft, The Netherlands, 2002.

[62] S. Kharrazi, "Steering based lateral performance control of long heavy vehicle combinations," PhD thesis, Chalmers University of Technology, 2012.

[63] N. Hata, S. Hasegawa, S. Takahashi, K. Ito, and T. Fujishiro, "A control method for 4WS truck to suppress excursion of a body rear overhang," Society of Automobile Engineers, Tech. Rep., 1989.

[64] I. Notsu, S. Takahashi, and Y. Watanabe, "Investigation into turning behavior of semi-trailer with additional trailer-wheel steering: A control method for trailer-wheel steering to minimize trailer rear-overhang swing in short turns," Society of Automobile Engineers, Tech. Rep., 1991.

[65] Q. Miao and D. Cebon, "Effects of operating off-highway on the tracking performance of a path-following steering system," in *FISITA World Automot. Congr.*, Maastricht, 2014.

[66] VSE, "Product information for ETS trailer," Veenendaal, The Netherlands, 2009. [Online]. Available: http://www.v-s-e.com/uploads/documents/leaflet-ets-trailer-en.pdf (Accessed 2017-06-05).

[67] "Product overview: mechanical and hydraulic steering systems and axle suspensions." [Online]. Available: http://jic.jost-world.com/static/upload/pdf/FLY/32004_FLY_TRIDEC_Uebersicht_Image_160824_EN_SCREEN.pdf

[68] Y. Lu, S. V. Byrne, P. A. VanOphem, and J. D. Harris, "Trailer angle detection system," 2014.

[69] S. J. Buckley, "Trailer detection system," 2010.

[70] T. Ehlgen, T. Pajdla, and D. Ammon, "Eliminating blind spots for assisted driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 4, pp. 657–665, 2008.

[71] J. Schikora, U. Berg, and D. Zöbel, "Berührungslose Winkelbestimmung zwischen Zugfahrzeug und Anhänger," in *Aktuelle Anwendungen Tech. und Wirtschaft*, ser. Informatik aktuell, W. Halang and P. Holleczek, Eds. Springer Berlin Heidelberg, 2009, pp. 11–20.

[72] L. Caup, J. Salmen, I. Muharemovic, and S. Houben, "Video-based trailer detection and articulation estimation," in *IEEE Intell. Veh. Symp.* Gold Coast: IEEE, Jun. 2013, pp. 1179–1184.

[73] M. P. Harris, "Application of computer vision systems to heavy goods vehicles: visual sensing of articulation angle," MPhil thesis, University of Cambridge, 2013.

[74] C. Fuchs, S. Eggert, B. Knopp, and D. Zobel, "Pose detection in truck and trailer combinations for advanced driver assistance systems," in *Proc. IEEE Intell. Veh. Symp.* Dearborn, MI: IEEE, Jun. 2014, pp. 1175–1180.

[75] C. Fuchs, F. Neuhaus, and D. Paulus, "Advanced 3-D trailer pose estimation for articulated vehicles," in *IEEE Intell. Veh. Symp.* Seoul: IEEE, 2015, pp. 211–216.

[76] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[77] Tangfei Tao, Ja Choon Koo, Hyouk-Ryeol Choi, T. Tao, J. C. Koo, and H. R. Choi, "A fast block matching algorthim for stereo correspondence," in *IEEE Conf. Cybern. Intell. Syst.* Chengdu: IEEE, Sep. 2008, pp. 38–41.

[78] L.-H. H. Lin, P. D. Lawrence, and R. Hall, "Robust outdoor stereo vision SLAM for heavy machine rotation sensing," *Mach. Vis. Appl.*, vol. 24, no. 1, pp. 205–226, Nov. 2013.

[79] U. Larsson, C. Zell, K. Hyyppa, and A. Wernersson, "Navigating an articulated vehicle and reversing with a trailer," in *Proc. IEEE Int. Conf. Robot. Autom.* IEEE Comput. Soc. Press, 1994, pp. 2398–2404.

[80] C. Cariou, R. Lenain, B. Thuilot, and P. Martinet, "Path following of a vehicle-trailer system in presence of sliding: Application to automatic guidance of a towed agricultural implement," in *Int. Conf. Intell. Robot. Syst.* Taipei: IEEE, 2010, pp. 4976–4981.

[81] D. M. Helmick, Y. Cheng, D. S. Clouse, L. H. Matthies, and S. I. Roumeliotis, "Path following using visual odometry for a Mars rover in high-slip environments," in *Aerosp. Conf.* IEEE, 2004, pp. 772–789.

[82] D. Scaramuzza and F. Fraundorfer, "Visual odometry part I - The first 30 years and fundamentals," *IEEE Robot. Autom. Mag.*, vol. 18, no. 4, pp. 80–92, 2011.

[83] "Autowatch AASDS side detection system." [Online]. Available: http://www.autowatch.co.uk/product/AASDS/AASDS.html (Accessed 2014-06-06).

[84] "Backwatch safety products: blindspot detection system." [Online]. Available: http://www.backwatch.co.uk/products/blindspot-detection-system/ (Accessed 2017-06-29).

[85] Y. Jia, "An automated cyclist collision avoidance system for Heavy Goods Vehicles," PhD thesis, University of Cambridge, 2014.

[86] Y. Jia and D. Cebon, "Field testing of a cyclist collision avoidance system for heavy goods vehicles," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 4359–4367, Jun. 2016.

[87] D. Kissinger, *Millimeter-wave receiver concepts for 77 GHz automotive radar in silicon-germanium technology*. Springer Science+Business Media, 2012.

[88] B. Schwarz, "Lidar: Mapping the world in 3D," *Nat. Photonics*, vol. 4, no. 7, pp. 429–430, 2010.

[89] C. Ilas, "Electronic sensing technologies for autonomous ground vehicles: A review," in *8th Int. Symp. Adv. Top. Electr. Eng.* IEEE, May 2013, pp. 1–6.

[90] Velodyne, "High definition lidar HDL-64E S2 - datasheet," 2010. [Online]. Available: http://velodynelidar.com/lidar/products/brochure/HDL-64EDataSheet.pdf (Accessed 2017-06-29).

[91] W. Knight, "Driverless cars are further away than you think," *MIT Technol. Rev.*, 2013.

[92] M. Bertozzi, A. Broggi, M. Cellario, A. Fascioli, P. Lombardi, and M. Porta, "Artificial vision in road vehicles," *Proc. IEEE*, vol. 90, no. 7, pp. 1258–1271, Jul. 2002.

[93] A. Discant, A. Rogozan, C. Rusu, and A. Bensrhair, "Sensors for obstacle detection - a survey," in *30th Int. Spring Semin. Electron. Technol.* Cluj-Napoca: IEEE, May 2007, pp. 100–105.

[94] E. D. Dickmanns, "The development of machine vision for road vehicles in the last decade," in *Intell. Veh. Symp. IEEE*, vol. 1, 2002, pp. 268–281.

[95] M. Sivak, "The information that drivers use: Is it indeed 90 percent visual?" *UMTRI Res. Rev.*, vol. 29, no. 1-9, 1998.

[96] H. Cho, P. E. Rybski, and W. Zhang, "Vision-based bicycle detection and tracking using a deformable part model and an EKF algorithm," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.* Madeira: IEEE, Jun. 2010, pp. 1875–1880.

[97] R. Gonzalez, F. Rodriguez, J. L. Guzman, C. Pradalier, and R. Siegwart, "Combined visual odometry and visual compass for off-road mobile robots localization," *Robotica*, vol. 30, no. 06, pp. 865–878, 2012.

[98] M. Schlipsing, J. Schepanek, and J. Salmen, "Video-based roll angle estimation for two-wheeled vehicles," in *Intell. Veh. Symp. (IV), IEEE*. Baden-Baden: IEEE, Jun. 2011, pp. 876–881.

[99] M. Schlipsing, J. Salmen, B. Lattke, K. G. Schroter, and H. Winner, "Roll angle estimation for motorcycles: comparing video and inertial sensor approaches," in *Intell. Veh. Symp. (IV), IEEE*, Alcala de Henares, 2012, pp. 500–505.

[100] I. F. Mondragón, P. Campoy, C. Martinez, and M. Olivares, "Omnidirectional vision applied to Unmanned Aerial Vehicles (UAVs) attitude and heading estimation," *Rob. Auton. Syst.*, vol. 58, no. 6, pp. 809–819, 2010.

[101] I. F. Mondragón, M. A. Olivares-Méndez, P. Campoy, C. Martínez, and L. Mejias, "Unmanned aerial vehicles UAVs attitude, height, motion estimation and control using visual systems," *Auton. Robots*, vol. 29, no. 1, pp. 17–34, 2010.

[102] A. E. R. Shabayek, C. Demonceaux, O. Morel, and D. Fofi, "Vision based UAV attitude estimation: progress and insights," *J. Intell. Robot. Syst.*, vol. 65, no. 1-4, pp. 295–308, Aug. 2011.

[103] Q. Miao, "Active steering off-road for long combination vehicles (first year report)," Ph.D. dissertation, University of Cambridge, Jul. 2012.

[104] Y. Wang, B. M. Nguyen, H. Fujimoto, and Y. Hori, "Multirate Estimation and Control of Body Slip Angle for Electric Vehicles Based on Onboard Vision System," *Ind. Electron. IEEE Trans.*, vol. 61, no. 2, pp. 1133–1143, 2014.

[105] Bobet and Schmid, "Obstacle detection analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 796–799.

[106] A. Broggi, C. Caraffi, R. Fedriga, and P. Grisleri, "Obstacle detection with stereo vision for off-road vehicle navigation," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work.*, vol. 3. San Diego, CA: IEEE, 2005, pp. 65–65.

[107] J.-C. C. Burie, J.-L. L. Bruyelle, and J.-G. G. Postaire, "Detecting and localizing obstacles in front of a moving vehicle using linear stereo vision," *Math. Comput. Model.*, vol. 22, no. 4-7, pp. 235–246, Aug. 1995.

[108] K. Huh, J. Park, J. Hwang, and D. Hong, "A stereo vision-based obstacle detection system in vehicles," *Opt. Lasers Eng.*, vol. 46, no. 2, pp. 168–178, 2008.

[109] R. Manduchi, A. Castano, A. Talukder, and L. Matthies, "Obstacle detection and terrain classification for autonomous off-road navigation," *Auton. Robots*, vol. 18, no. 1, pp. 81–102, 2005.

[110] Z. Zhang, R. Weiss, and A. R. Hanson, "Qualitative obstacle detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 554–559.

[111] "Volvo Car Group introduces world-first cyclist detection with full auto brake," 2013. [Online]. Available: https://www.media.volvocars.com/global/en-gb/media/pressreleases/48277 (Accessed 2017-06-29).

[112] "Fusion Processing: CycleEye." [Online]. Available: http://www.fusionproc.com/products/ (Accessed 2017-06-29).

[113] R. C. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty," *Int. J. Rob. Res.*, vol. 5, no. 4, pp. 56–68, Dec. 1986.

[114] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in *Proc. 8th Natl. Conf. Artif. Intell. Conf. Innov. Appl. Artif. Intell.*, vol. 68, no. 2. Edmonton, Alberta: AAAI, 2002, pp. 593–598.

[115] M. Montemerlo, S. Thrun, D. Roller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," *Proc. Sixt. Int. Jt. Conf. Artif. Intell.*, pp. 1151–1156, 2003.

[116] G. Grisetti, R. Kummerle, C. Stachniss, and W. Burgard, "A tutorial on Graph-Based SLAM," *IEEE Intell. Transp. Syst. Mag.*, vol. 2, no. 4, pp. 31–43, 2010.

[117] A. Angeli, S. Doncieux, J.-A. Meyer, and D. Filliat, "Visual topological SLAM and global localization," in *2009 IEEE Int. Conf. Robot. Autom.* IEEE, May 2009, pp. 4300–4305.

[118] Z. Lu, Z. Hu, and K. Uchimura, "SLAM estimation in dynamic outdoor environments: A review," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5928 LNAI, pp. 255–267, 2009.

[119] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: real-time single camera SLAM." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–67, Jun. 2007.

[120] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

[121] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *arXiv Prepr.*, Oct. 2016.

[122] J. Engel, T. Sch, and D. Cremers, "LSD-SLAM: Large-Scale Direct monocular SLAM," in *Proc. 13th Eur. Conf. Comput. Vis.*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., no. Part 2. Zurich: Springer, 2014, pp. 834–849.

[123] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," *6th IEEE ACM Int. Symp. Mix. Augment. Real.*, Nov. 2007.

[124] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle Adjustment — A Modern Synthesis," in *Vis. Algorithms Theory Pract.* Springer Berlin Heidelberg, 2000, vol. 1883, pp. 298–372.

[125] T. Pire, T. Fischer, J. Civera, P. De Cristoforis, and J. J. Berlles, "Stereo parallel tracking and mapping for robot localization," in *IEEE Int. Conf. Intell. Robot. Syst.* IEEE/RSJ, 2015, pp. 1373–1378.

[126] "Parallel Tracking And Mapping for small AR workspaces: Licence and source code download page," 2008. [Online]. Available: http://www.robots.ox.ac.uk/~gk/PTAM/download.html (Accessed 2017-07-04).

[127] "S-PTAM: Stereo Parallel Tracking and Mapping (GitHub)." [Online]. Available: https://github.com/lrse/sptam (Accessed 2017-07-04).

[128] A. Kendall, M. Grimes, and R. Cipolla, "PoseNet: A convolutional network for real-time 6-dof camera relocalization," in *Proc. IEEE Int. Conf. Comput. Vis.* IEEE, 2016, pp. 2938–2946.

[129] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.

[130] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of Kinect depth data for indoor mapping applications," *Sensors (Basel)*, vol. 12, no. 2, pp. 1437–54, 2012.

[131] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "The KITTI Vision Benchmark Suite: Visual Odometry / SLAM Evaluation 2012." [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_odometry.php (Accessed 2017-06-06).

[132] I. Cvisic and I. Petrovic, "Stereo odometry based on careful feature selection and tracking," in *2015 Eur. Conf. Mob. Robot.* IEEE, Sep. 2015, pp. 1–6.

[133] A. Geiger, J. Ziegler, and C. Stiller, "StereoScan: Dense 3d reconstruction in real-time," in *IEEE Intell. Veh. Symp.* Baden-Baden: IEEE, Jun. 2011, pp. 963–968.

[134] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect Visual Odometry for monocular and multicamera systems," *IEEE Trans. Robot.*, vol. 33, no. 2, pp. 249–265, Apr. 2017.

[135] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab, "Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7724 LNCS, no. PART 1, pp. 548–562, 2013.

[136] R. Opromolla, G. Fasano, G. Rufino, and M. Grassi, "A model-based 3D template matching technique for pose acquisition of an uncooperative space object," *Sensors (Switzerland)*, vol. 15, no. 3, pp. 6360–6382, 2015.

[137] S. Holzer, S. Hinterstoisser, S. Ilic, and N. Navab, "Distance transform templates for object detection and pose estimation," in *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009*, 2009, pp. 1177–1184.

[138] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[139] J. P. Lewis, "Fast normalized cross-correlation," *Vis. Interface*, vol. 10, no. 1, 1995.

[140] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey." *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[141] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR 2000)*, vol. 2, no. 8, pp. 142–149, 2000.

[142] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[143] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[144] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[145] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1653–1660.

[146] C. C. de Saxe and D. Cebon, "A visual template-matching method for articulation angle measurement," in *Proc. 18th IEEE Int. Conf. Intell. Transp. Syst.* Las Palmas: IEEE, 2015, pp. 626–631.

[147] C. C. de Saxe and D. Cebon, "Visual odometry for trailer off-tracking estimation," in *Proc. 14th Int. Symp. Heavy Veh. Transp. Technol.*, J. J. de Pont, Ed. Rotorua, NZ: TERNZ Ltd., 2016, pp. 138–149.

[148] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. London. Ser. B, Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, Feb. 1980.

[149] C. Harris, "Geometry from visual motion," in *Act. Vis.*, A. Blake and A. Yuille, Eds. Cambridge, MA: MIT Press, 1992, pp. 263–284.

[150] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[151] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention," *Int. J. Comput. Vis.*, vol. 11, no. 3, pp. 283–318, 1993.

[152] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Comput. Vis. – ECCV 2006 9th Eur. Conf. Comput. Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I*, A. Leonardis, H. Bischof, and A. Pinz, Eds. Springer Berlin Heidelberg, 2006, pp. 430–443.

[153] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. San Diego, CA: IEEE, 2005, pp. 886–893.

[154] K. Mikolajczyk and C. Schmid, "Performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–30, Oct. 2005.

[155] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, May 2001.

[156] D. J. Fleet and Y. Weiss, "Optical flow estimation," in *Handb. Math. Model. Comput. Vis.*, N. Paragios, Y. Chen, and O. Faugeras, Eds. Elsevier, 2006, pp. 237 – 257.

[157] R. Hartley and A. Zisserman, *Multiple View Geometry*, 2nd ed. Cambridge, UK: Cambridge University Press, 2003.

[158] R. Cipolla, "Computer vision and robotics: projection (4F12 lecture notes, Cambridge University Engineering Department)," 2013.

[159] J. Heikkilä and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Juan, 1997, pp. 1106–1112.

[160] F. Devernay and O. Faugeras, "Straight lines have to be straight," *Mach. Vis. Appl.*, vol. 13, no. 1, pp. 14–24, 2001.

[161] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 22, no. 11, pp. 1330–1334, 2000.

[162] "MATLAB Computer Vision System Toolbox Release 2017b," Natick, Massachusetts, USA.

[163] J.-Y. Bouguet, "Camera calibration toolbox for Matlab." [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html#ref (Accessed 2017-06-27).

[164] G. Bradski, "The OpenCV library," 2000. [Online]. Available: http://opencv.org/ (Accessed 2017-07-07).

[165] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Mach. Vis. Appl.*, vol. 12, no. 1, pp. 16–22, Jul. 2000.

[166] "MATLAB R2013b," Natick, MA, 2013. [Online]. Available: http://www.mathworks.co.uk/products/matlab/

[167] "Inventor Professional 2013," San Rafael, CA, 2013. [Online]. Available: http://www.autodesk.com/products/inventor/overview

[168] Y. Ma, S. Soatto, J. Kosecka, S. S. Sastry, J. Košecká, and S. S. Sastry, "Reconstruction from two calibrated views," in *An Invit. to 3-D Vis. from images to Geom. Model.*, 1st ed., S. S. Antman, J. E. Marsden, L. Sirovich, and S. Wiggins, Eds. New York: Springer Science+Business Media, 2004, ch. 5, pp. 109–170.

[169] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," 2008. [Online]. Available: http://www.vlfeat.org/

[170] A. Sibiryakov, "Fast and high-performance template matching method," in *Comput. Vis. Pattern Recognit.* IEEE, Jun. 2011, pp. 1417–1424.

[171] G. Papamakarios, G. Rizos, N. P. Pitsianis, and X. Sun, "Fast computation of local correlation coefficients on graphics processing units," in *Proc. SPIE*, vol. 7444, 2009.

[172] C. F. Olson and D. P. Huttenlocher, "Automatic target recognition by matching oriented edge pixels." *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 103–13, Jan. 1997.

[173] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1465–1479, Sep. 2006.

[174] O. Arslan, B. Demirci, H. Altun, and N. S. Tunaboylu, "A novel rotation-invariant template matching based on HOG and AMDF for industrial laser cutting applications," in *9th Int. Symp. Mechatronics its Appl.* IEEE, Apr. 2013, pp. 1–5.

[175] C. Zauner, "Implementation and benchmarking of perceptual image hash functions," Diplomarbeit, University of Applied Sciences Hagenberg, 2010.

[176] T. Dekel, S. Oron, M. Rubinstein, S. Avidan, and W. T. Freeman, "Best-Buddies Similarity for robust template matching," in *IEEE Conf. Comput. Vis. Pattern Recognit.* IEEE, Jun. 2015, pp. 2021–2029.

[177] D. E. Newland, *An introduction to random vibrations and spectral analysis*, 2nd ed. New York: Longman Group Limited, 1984.

[178] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[179] C. B. Winkler and J. Aurell, "Analysis and testing of the steady-state turning of multiaxle trucks," in *5th Int. Symp. Heavy Veh. Weight. Dimens.*, Maroochydore, Queensland, Australia, 1998, pp. 135–161.

[180] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman Filter to nonlinear systems," in *SPIE 3068, Signal Process. Sens. Fusion, Target Recognit. VI*, I. Kadar, Ed., vol. 3068. Orlando, FL: International Society for Optics and Photonics, Jul. 1997, pp. 182–193.

[181] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Adapt. Syst. Signal Process. Commun. Control Symp.*, vol. v. Lake Louise, Alta.: IEEE, 2000, pp. 153–158.

[182] E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–19, Jan. 2008.

[183] A. W. Bridgewater, "Analysis of second and third order steady-state tracking filters," in *AGARD Conf. Proc. no. 252 Strateg. Autom. Track Initiat.*, S. J. Rabinowitz, Ed. Monterey, CA: AGARD, Jun. 1979.

[184] H. Stewénius, C. Engels, and D. Nistér, "Recent developments on direct relative orientation," *ISPRS J. Photogramm. Remote Sens.*, vol. 60, no. 4, pp. 284–294, Jun. 2006.

[185] S. M. La Valle, "Geometric representations and transformations," in *Plan. algorithms*, 1st ed. Cambridge University Press, 2006, ch. 3, pp. 81–126.

[186] G. Sibley, C. Mei, I. Reid, and P. Newman, "Vast-scale outdoor navigation using adaptive relative bundle adjustment," *Int. J. Rob. Res.*, vol. 29, no. 8, pp. 958–980, 2010.

[187] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme," in *IEEE Intell. Veh. Symp.* San Diego, CA: IEEE, 2010, pp. 486–492.

[188] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Conf. Comput. Vis. Pattern Recognit.* Providence, RI: IEEE, Jun. 2012, pp. 3354–3361.

[189] C. F. Olson and H. Abi-Rached, "Wide-baseline stereo vision for terrain mapping," *Mach. Vis. Appl.*, vol. 21, no. 5, pp. 713–725, 2010.

[190] K. Konolige and M. Agrawal, "FrameSLAM: from bundle adjustment to real-time visual mapping," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 1066–1077, Oct. 2008.

[191] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, Sep. 2004.

[192] "AutoIt v3." [Online]. Available: https://www.autoitscript.com/site/autoit/ (Accessed 2017-07-13).

[193] "Volvo truck accessories - Volvo FM," 2014. [Online]. Available: http://www.volvotrucks.com.au/content/dam/volvo/volvo-trucks/masters/euro-6/pdf/trucks/volvo-fm/accessories/Volvo_Accessories_FM.pdf (Accessed 2017-06-09).

[194] "Orlaco All Time Vision Camera (ATVC) set - Self cleaning camera," 2017. [Online]. Available: https://www.orlaco.com/product/orlaco-all-time-vision-camera-atvc-self-cleaning-camera (Accessed 2017-06-09).

[195] "Continental ProViu: maneuvering without risk," 2012. [Online]. Available: https://www.automotiveworld.com/news-releases/continental-proviu-maneuvering-without-risk/ (Accessed 2017-06-29).

[196] "360° bird's eye view camera." [Online]. Available: http://www.autowatch.co.uk/product/DVR/BEV360.html (Accessed 2014-03-09).

[197] "Safety Shield Systems: introducing Cycle Safety Shield." [Online]. Available: http://safetyshieldsystems.com/cycle-safety-shield/ (Accessed 2017-06-29).

[198] Y.-L. Chang, L.-Y. Hsu, and O. T.-C. Chen, "Auto-calibration around-view monitoring system," in *78th IEEE Veh. Technol. Conf. (VTC Fall)*. IEEE, Sep. 2013, pp. 1–5.

[199] S. Lee and H. Shim, "Skewed stereo time-of-flight camera for translucent object imaging," *Image Vis. Comput.*, vol. 43, pp. 27–38, Nov. 2015.

Appendix A

Computer specifications

Table A.1 'Desktop computer' specifications (Windows)

Processor	Intel® Core™ i7-3930K CPU @ 3.20 GHz (6 cores, 12 threads)
Memory	8 GB DDR3 @ 800 MHz
Primary storage	Optical hard drive
Graphics	NVIDEA GeForce GTX 650, 1GB DDR5
Operating system	Windows® 7 Professional, 64-bit (Service Pack 1)
Software	MATLAB® 2013b

Table A.2 'Laptop computer' specifications (Linux)

Processor	Intel® Core™ i3-3110M CPU @ 2.4 GHz (2 cores, 4 threads)
Memory	4 GB DDR3 @ 1600 MHz
Primary storage	Optical hard drive
Graphics	Intel® HD Graphics 4000
Operating system	Ubuntu 14.04 LTS, 64-bit
Software	Python 2.7.6, OpenCV 2.4.8

Table A.3 'Linux computer' specifications

Processor	Intel® Core™ i7-4770 CPU @ 3.4 GHz (4 cores, 8 threads)
Memory	16 GB DDR3 @ 1600 MHz
Primary storage	Solid-state hard drive
Graphics	Intel® HD Graphics 4600
Operating system	Ubuntu 14.04 LTS, 64-bit
Software	Python 2.7.10, OpenCV 2.4.12

Appendix B

Unscented Kalman Filter

This is an overview of the theory of the Unscented Kalman Filter, following that presented in [180] and [181]. For a random variable $\mathbf{x}$ of dimension L with mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}_{\mathbf{x}}$, the $2L+1$ sigma points are calculated as follows ($i \in 1,...,L$):

$$\begin{aligned} \mathcal{X}_0 &= \bar{\mathbf{x}} \\ \mathcal{X}_i &= \bar{\mathbf{x}} + \left(\sqrt{(L+\lambda)\mathbf{P}_{\mathbf{x}}}\right)_i \\ \mathcal{X}_{i+n} &= \bar{\mathbf{x}} - \left(\sqrt{(L+\lambda)\mathbf{P}_{\mathbf{x}}}\right)_i \end{aligned} \tag{B.1}$$

The weights associated with each of these sigma points are:

$$\begin{aligned} W_0^{(m)} &= \lambda/(L+\lambda) \\ W_0^{(c)} &= \lambda/(L+\lambda) + (1-\alpha^2+\beta) \\ W_i^{(m)} &= W_i^{(c)} = 1/2(L+\lambda) \\ W_{i+n}^{(m)} &= W_{i+n}^{(c)} = 1/2(L+\lambda) \end{aligned} \tag{B.2}$$

where (c) and (m) denote applicability to the variable mean and covariance respectively (see later). $\lambda = \alpha^2(L+\kappa) - L$ is a scaling parameter. $\alpha = 0.001$, $\kappa = 0$, and $\beta = 2$ as per the recommended values in [181] for second order accuracy.

If $\mathbf{x}$ is a general state variable with assumed Gaussian statistics, estimated mean $\hat{\mathbf{x}}$ and covariance $\mathbf{P}$, and if process and measurement noise are $\mathbf{v}$ and $\mathbf{u}$ respectively, then for time steps $k \in \{1,...,\infty\}$ the sigma points are:

$$\boldsymbol{\mathcal{X}}_{k-1}^a = \left[\hat{\mathbf{x}}_{k-1}^a \quad \hat{\mathbf{x}}_{k-1}^a \pm \sqrt{(L+\lambda)\mathbf{P}_{k-1}^a}\right] \tag{B.3}$$

Here the state variable and its covariance have been augmented with the noise variables so that $\mathbf{x}^a = [\mathbf{x}^\top \mathbf{v}^\top \mathbf{n}^\top]^\top$ and $\mathbf{P}^a = \text{diag}(\mathbf{P}, \mathbf{P_v}, \mathbf{P_n})$. The resulting sigma points are therefore in the form $\boldsymbol{\mathcal{X}}^a = [(\boldsymbol{\mathcal{X}}^x)^\top (\boldsymbol{\mathcal{X}}^v)^\top (\boldsymbol{\mathcal{X}}^n)^\top]^\top$.

What remains is the familiar Kalman Filter time update and measurement update sequence. The time update is performed by first passing the sigma points through the process model, **F**:

$$\boldsymbol{\mathcal{X}}^x_{k|k-1} = \mathbf{F}[\boldsymbol{\mathcal{X}}^x_{k-1}, \boldsymbol{\mathcal{X}}^v_{k-1}] \tag{B.4}$$

The posterior mean and covariance of the state are then calculated:

$$\hat{\mathbf{x}}^-_k = \sum_{i=0}^{2L} W_i^{(m)} \mathcal{X}^x_{i,k|k-1} \tag{B.5}$$

$$\mathbf{P}^-_k = \sum_{i=0}^{2L} W_i^{(c)} [\mathcal{X}^x_{i,k|k-1} - \hat{\mathbf{x}}^-_k][\mathcal{X}^x_{i,k|k-1} - \hat{\mathbf{x}}^-_k]^\top \tag{B.6}$$

Finally, if **H** is the observation model, the posterior mean of the observation may be found as follows:

$$\boldsymbol{\mathcal{Y}}^x_{k|k-1} = \mathbf{H}[\boldsymbol{\mathcal{X}}^x_{k|k-1}, \boldsymbol{\mathcal{X}}^n_{k|k-1}] \tag{B.7}$$

$$\hat{\mathbf{y}}^-_k = \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}_{i,k|k-1} \tag{B.8}$$

The measurement update is performed by first calculating the innovation covariance, $\mathbf{P_{yy}}$, and cross correlation matrix, $\mathbf{P_{xy}}$:

$$\mathbf{P}_{\mathbf{yy},k} = \sum_{i=0}^{2L} W_i^{(c)} [\mathcal{Y}_{i,k|k-1} - \hat{\mathbf{y}}^-_k][\mathcal{Y}_{i,k|k-1} - \hat{\mathbf{y}}^-_k] \tag{B.9}$$

$$\mathbf{P}_{\mathbf{xy},k} = \sum_{i=0}^{2L} W_i^{(c)} [\mathcal{X}_{i,k|k-1} - \hat{\mathbf{x}}^-_k][\mathcal{Y}_{i,k|k-1} - \hat{\mathbf{y}}^-_k] \tag{B.10}$$

Finally, the estimated state and its covariance may be calculated as follows:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}^-_k + \mathcal{K}(\mathbf{y}_k - \hat{\mathbf{y}}^-_k) \tag{B.11}$$

$$\mathbf{P}_k = \mathbf{P}^-_k - \mathcal{K}\mathbf{P}_{\mathbf{yy},k}\mathcal{K}^\top \tag{B.12}$$

where

$$\mathcal{K} = \mathbf{P}_{\mathbf{xy},k}\mathbf{P}^{-1}_{\mathbf{yy},k} \tag{B.13}$$

Appendix C

Camera calibration results

Chapter 4, single camera

Template matching (using the Camera Calibration Toolbox for MATLAB):

$$(f_x, f_y) = (393.8, 395.7) \text{ pixels}$$
$$(u_0, v_0) = (328.4, 247.3) \text{ pixels}$$
$$(k_1, k_2, p_1, p_2, k_3) = (-0.3013, 0.0751, 0.0028, 0.00044, 0)$$

PTAM (using the PTAM camera calibration module):

$$(f_x, f_y) = (355.9, 355.5) \text{ pixels}$$
$$(u_0, v_0) = (333.8, 248.5) \text{ pixels}$$
$$s = 0.9813$$

Chapter 5, single camera

PTAM (using the PTAM camera calibration module):

$$(f_x, f_y) = (357.7, 358.8) \text{ pixels}$$
$$(u_0, v_0) = (329.5, 249.3) \text{ pixels}$$
$$s = 0.9821$$

Chapter 7, stereo cameras

Visual odometry (using the OpenCV calibration functions for C++):

$$\mathrm{K}_l = \begin{bmatrix} 886.8 & 0 & 631.4 \\ 0 & 887.0 & 539.0 \\ 0 & 0 & 1 \end{bmatrix} \text{ pixels}$$

$$(k_1,\ k_2,\ k_3,\ k_4,\ k_5,\ k_6)_l = (-0.1899,\ 0.5640,\ -0.0009,\ -0.0012,\ 0.2202,\ 0.2790)$$

$$(p_1,\ p_2)_l = (0.3351,\ 0.5668)$$

$$\mathrm{K}_r = \begin{bmatrix} 885.0 & 0 & 662.1 \\ 0 & 886.0 & 545.4 \\ 0 & 0 & 1 \end{bmatrix} \text{ pixels}$$

$$(k_1,\ k_2,\ k_3,\ k_4,\ k_5,\ k_6)_r = (0.5792,\ 0.0215,\ -0.0022,\ 0.0000,\ 0.0047,\ 1.0460)$$

$$(p_1,\ p_2)_r = (0.1651,\ 0.0062)$$

$$\mathrm{R}_{rl,500\mathrm{mm}} = \begin{bmatrix} 0.9997 & -0.0021 & -0.02544 \\ 0.0018 & 0.9999 & -0.01150 \\ 0.0255 & 0.0115 & 1.0000 \end{bmatrix}$$

$$\mathbf{T}_{rl,500\mathrm{mm}} = \begin{bmatrix} -49.9680,\ -0.0535,\ -0.0645 \end{bmatrix}^{\mathsf{T}} \text{ cm}$$

$$\mathrm{R}_{rl,700\mathrm{mm}} = \begin{bmatrix} 0.9996 & 0.0036 & -0.0283 \\ -0.0040 & 0.9999 & -0.0117 \\ 0.0283 & 0.0118 & 0.9995 \end{bmatrix}$$

$$\mathbf{T}_{rl,700\mathrm{mm}} = \begin{bmatrix} -70.0820,\ 0.1262,\ -0.2792 \end{bmatrix}^{\mathsf{T}} \text{ cm}$$

$$\mathrm{R}_{rl,900\mathrm{mm}} = \begin{bmatrix} 0.9997 & 0.0009 & -0.0260 \\ -0.0013 & 0.9999 & -0.0154 \\ 0.0260 & 0.0154 & 0.9995 \end{bmatrix}$$

$$\mathbf{T}_{rl,900\mathrm{mm}} = \begin{bmatrix} -89.8847,\ 0.0798,\ -0.5974 \end{bmatrix}^{\mathsf{T}} \text{ cm}$$